

OM of: US-09-303-518d-131 to: SwissProt\_40:\* out\_format : pfs

Date: Jun 30, 2002 8:35 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framer\_n2p.model -DEV=xlh  
-O=/cn2\_1/USPTO\_spo1/US09303518/runat\_28062002\_142714\_4362/app\_query.fasta\_1.23501  
-DB=SwissProt\_40 -OFTM=fastan -SUFFIX=rsp -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -LOOPEXT=0.000  
-DELOP=6.000 -DELEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500  
-MATRIX=blosum62  
-TRANS=human40.cdi -LIST=100 -DOALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09303518 -CGN1\_1\_440 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-131

Query length: 1344

Database: SwissProt\_40.\*

Database sequences: 105224

Database length: 38719550

Search time (sec): 217.960000

score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
SwissProt_40:NQRA_NEIMA	2222.00	2930.16	3.6e-156	447	! Q9jv88 neisseria meningitidis
SwissProt_40:NQRA_NEIMB	2221.00	2928.83	4.3e-156	447	! Q9k0m3 neisseria meningitidis
SwissProt_40:NQRA_PASMU	1770.50	2240.55	9.4e-118	446	! O8clb1 pasteurella multocida
SwissProt_40:NQRA_HAEIN	1646.00	2168.46	9.7e-114	447	! P43955 haemophilus influenza
SwissProt_40:NQRA_VIBHA	1491.50	1964.17	2.3e-102	446	! Q9fvl1 vibrio harveyi. na(+)
SwissProt_40:NQRA_VIBAL	1488.50	1960.20	3.9e-102	446	! Q9fvl1 vibrio harveyi. na(+)
SwissProt_40:NQRA_VIBCH	1476.50	1944.33	3.0e-101	446	! Q9fvl1 vibrio cholerae. na(+)
SwissProt_40:NQRA_PSEAE	1358.00	1787.65	1.6e-92	445	! Q9kz6 pseudomonas aeruginos
SwissProt_40:NQRA_CHLNU	453.00	590.48	7.3e-26	465	! Q9plu3 chlamydia muridarum
SwissProt_40:NQRA_CHLNU	450.50	587.13	1.1e-25	467	! Q9z722 chlamydia pneumoniae
SwissProt_40:NQRA_CHLNU	432.00	562.71	2.6e-24	465	! O84639 chlamydia trachomatis
SwissProt_40:NQRA_VIBCH	172.00	214.17	4.0e-05	774	! Q9kz88 vibrio cholerae. elec
SwissProt_40:NQRA_YN96	143.00	172.23	0.0059	1140	! Q9kz88 vibrio cholerae. elec
SwissProt_40:NQRA_VIBCH	135.50	165.38	0.0199	819	! P71397 haemophilus cerevis
SwissProt_40:NQRA_PSEAE	132.00	161.27	0.0356	774	! Q9hyb8 pseudomonas aeruginos
SwissProt_40:NQRA_BUCAI	131.00	164.51	0.0385	473	! P57215 buchnera aphidicola
SwissProt_40:N121_RAT	129.50	153.91	0.0590	1199	! P52591 rattus norvegicus (rat
SwissProt_40:HKR1_YEAST	127.00	146.83	0.0974	1802	! P41809 saccharomyces cerevi
SwissProt_40:RNFC_BHOC	126.50	157.70	0.0839	519	! Q52716 rhodobacter capsulatu
SwissProt_40:RNFC_ECO57	121.00	147.14	0.2280	740	! P58324 escherichia coli o157
SwissProt_40:RNFC_ECOLI	117.00	141.85	0.4494	740	! P77611 escherichia coli. ele
SwissProt_40:OPD2_MYCOA	116.50	146.05	0.4433	438	! O49110 mycoplasma capricolum
SwissProt_40:FAS1_SCHPO	114.50	129.01	0.8330	2073	! Q9uug0 s fatty acid synthas
SwissProt_40:PPSA_PYRAB	113.00	135.62	0.9026	819	! Q9v2h7 pyrococcus abyssi. p
SwissProt_40:CHLN_SYNP7	112.50	140.18	0.8838	466	! P54208 synecococcus sp. (st
SwissProt_40:PNR8_MOUSE	112.00	134.49	1.07	802	! P29352 mus musculus (mouse)
SwissProt_40:PPSA_PYRAB	111.00	133.00	1.27	817	! P42850 pyrococcus furiosus
SwissProt_40:HEFC1_MESAU	109.50	122.32	1.95	1299	! P51611 mesocricetus auratus
SwissProt_40:N121_HUMAN	108.50	125.91	2.09	1209	! Q9v2n3 homo sapiens (human)
SwissProt_40:MSB2_YEAST	107.00	123.27	2.73	1306	! P32334 saccharomyces cerevi
SwissProt_40:AMVH_YEAST	107.00	122.94	2.75	1367	! P08640 saccharomyces cerevi
SwissProt_40:MUC1_MOUSE	106.50	128.46	2.59	630	! Q02496 mus musculus (mouse)
SwissProt_40:ATPB_NICRU	105.00	129.65	3.19	498	! P26530 nicotiana rustica (az
SwissProt_40:ATPB_TORAC	105.00	129.65	3.19	498	! P00826 nicotiana tabacum (co
SwissProt_40:PPSA_PYRHO	105.00	125.02	3.51	821	! O57830 pyrococcus horikoshii
SwissProt_40:SMB2_MESAU	105.00	123.30	3.63	989	! Q06560 mesocricetus auratus
SwissProt_40:ILVE_BACSU	104.50	131.94	3.27	362	! P39376 bacillus subtilis. pu
SwissProt_40:MUC1_MESAU	104.50	126.16	3.68	676	! Q06528 mesocricetus auratus
SwissProt_40:MURF_HAEIN	104.00	129.12	3.72	457	! P45061 haemophilus influenza
SwissProt_40:SC18_SCHPO	103.50	123.37	4.49	792	! Q9p7q4 schizosaccharomyces p

SwissProt_40:VGP3_EBV	103.50	122.12	4.61	907	! P03200 epstein-barr virus (
SwissProt_40:TEGU_HCMVA	103.50	113.74	5.46	2241	! P16785 human cytomegalov
SwissProt_40:SGS3_DROER	102.00	129.55	4.91	328	! P13730 nicotiana erecta
SwissProt_40:ATPB_NICPL	102.00	125.68	5.31	498	! P26529 schizosaccharomyces
SwissProt_40:HUTH_BACSU	101.50	124.84	5.80	508	! P10944 bacillus subtilis.
SwissProt_40:IF2G_YEAST	101.50	124.50	5.84	527	! P32481 saccharomyces cere
SwissProt_40:APMU_PIG	101.50	117.27	6.76	1150	! P12021 sus scrofa (pig). a
SwissProt_40:RAUO_METUA	101.00	127.95	5.85	338	! P54049 methanococcus jann
SwissProt_40:YN96_YEAST	101.00	116.88	7.32	1117	! P53753 saccharomyces cer
SwissProt_40:DTXH_CORBE	100.50	122.61	7.00	560	! P00589 corynebacter beta.
SwissProt_40:CHEB_PYRHO	100.00	125.84	7.04	368	! O58192 pyrococcus horikos
SwissProt_40:FTG2_YEAST	100.00	112.18	9.29	1609	! P25653 saccharomyces cer
SwissProt_40:ATPB_PYLLI	99.50	122.70	8.06	481	! P26532 pyllaella littoral
SwissProt_40:RNFC_PASMU	99.50	117.59	8.94	835	! Q9cnp2 pasteurella multoc
SwissProt_40:PPSA_AERPE	99.50	117.48	8.96	845	! Q9yec5 aeropyrum pernix.
SwissProt_40:RPOB_RICPR	99.50	112.98	9.82	1374	! O52271 rickettsia prowaz
SwissProt_40:DTXK_EMENI	99.50	102.32	12.18	4344	! P45444 ericella nidula
SwissProt_40:DTXK_COROM	98.50	119.97	9.83	560	! P00587 corynebacter omega.
SwissProt_40:DTXK_CORBE	98.50	119.85	9.85	567	! P00588 corynebacter beta. d
SwissProt_40:MKC7_YEAST	98.50	119.39	9.95	596	! P53379 saccharomyces cere
SwissProt_40:MUE_DROME	98.00	111.56	12.52	1293	! P24785 drosophila melanog
SwissProt_40:MURA_ECOLI	97.50	121.33	11.03	419	! P28909 escherichia coli.
SwissProt_40:VL2_HPV70	97.50	120.35	11.25	466	! P50801 human papillomavir
SwissProt_40:CHLN_SYNP3	97.50	120.29	11.27	469	! P28372 synecocystis sp.
SwissProt_40:URE2_HELFE	97.50	118.50	11.68	569	! Q08716 helicobacter felis
SwissProt_40:YN26_MYCTU	97.00	115.96	13.21	697	! P71886 mycobacterium tube
SwissProt_40:DPO1_THEAQ	97.00	114.32	13.21	832	! P19821 thermus aquaticus
SwissProt_40:LHS1_YEAST	97.00	113.79	13.80	881	! P36016 saccharomyces cere
SwissProt_40:FAS1_CANAL	97.00	106.03	16.15	2037	! P34731 c fatty acid synt
SwissProt_40:SRF2_BACSU	96.50	100.13	19.55	3587	! Q04747 bacillus subtilis
SwissProt_40:YG1F_YEAST	96.00	116.81	14.98	551	! P45314 saccharomyces cere
SwissProt_40:ANP_NOTCO	96.00	113.48	16.02	790	! P24856 notothenia coriicep
SwissProt_40:SMB2_MOUSE	96.00	111.36	16.72	993	! P40694 mus musculus (mous
SwissProt_40:SON_HUMAN	96.00	103.09	19.78	2426	! P18583 homo sapiens (huma
SwissProt_40:INDH_PYRAB	95.50	117.33	15.91	485	! Q9uy49 pyrococcus abyssi
SwissProt_40:NSF_CRIGR	95.50	113.37	17.25	744	! P18708 cricetulus griseus
SwissProt_40:YA54_METJA	95.50	111.66	17.85	895	! Q58454 methanococcus jann
SwissProt_40:BI0B_BACSH	95.00	120.18	16.13	332	! P19206 bacillus anthracis
SwissProt_40:MS11_AGRRH	95.00	117.79	16.94	430	! P27874 agrobacterium rhiz
SwissProt_40:ATPB_DICDH	95.00	116.75	17.30	481	! P30158 dictyocha dichotoma
SwissProt_40:ADEC_METH	95.00	115.71	17.66	538	! Q28952 methanobacterium t
SwissProt_40:AGAL_YEAST	95.00	112.95	18.68	725	! P32323 saccharomyces cere
SwissProt_40:NSF_HUMAN	95.00	103.39	22.67	2035	! P51610 mus musculus (hum
SwissProt_40:ICP0_HSVB	94.50	112.05	20.43	744	! P46460 mus musculus (mous
SwissProt_40:GLYA_LACLA	94.00	116.79	19.93	415	! Q9ch77 lactococcus lactis
SwissProt_40:ATPB_ANSP	94.00	115.41	20.50	482	! P06540 anabaena sp. (stra
SwissProt_40:ATPB_NICSP	94.00	115.10	20.53	498	! P26540 anabaena sp. (cob
SwissProt_40:FGMP_PEA	94.00	112.99	21.53	626	! Q9sm59 pisum sativum (garde
SwissProt_40:ICP0_HSVB	94.00	112.84	21.60	636	! P53882 saccharomyces cere
SwissProt_40:ICP0_HSVB	94.00	112.27	21.84	676	! P29128 bovine herpesvirus
SwissProt_40:MID2_YEAST	93.50	117.04	21.30	376	! P36027 saccharomyces cere
SwissProt_40:IF2G_SCHPO	93.50	115.46	21.99	446	! Q09130 schizosaccharomyce
SwissProt_40:AMPH_CHICK	93.50	111.53	23.82	582	! P50478 gallus gallus (chi
SwissProt_40:CTPV_MYCTU	93.50	110.41	24.37	770	! P77894 mycobacterium tube
SwissProt_40:EDD_ZYMOA	93.00	112.32	25.17	583	! P21909 zymomonas mobilis.
SwissProt_40:FSHR_MACFA	93.00	110.70	26.02	695	! P32212 macoma fasciculari
SwissProt_40:CARB_PSEAE	93.00	106.68	28.22	1072	! P38100 pseudomonas aerug
SwissProt_40:ALSI_CANAL	93.00	105.19	29.09	1260	! P46590 candida albicans
SwissProt_40:VGLX_HSVB	92.50	108.77	29.06	797	! P28968 equine herpesvirus
SwissProt_40:VP6_BTWIS	92.00	116.33	26.78	328	! Q03328 bluetongue virus (s

seq\_name: SwissProt\_40:NQRA\_NEIMA

seq\_documentation\_block:

ID	NQRA_NEIMA	STANDARD;	PRT;	447 AA.
AC	Q9JVP8;			
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DE	Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)			
DE	(Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex subunit A) (NQR-1 subunit A).			
GN	NQRA OR NQA0752.			

```
OS Neisseria meningitidis (serogroup A),
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holford S.,
RA Jagels K., Leather S., Moule K., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRE, NQRF,
CC AND NORF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL162754; CAB84035.1; -.
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 447 AA; 48673 MW; B808CAA82C901AID CRC64;

alignment_scores:
  Quality: 2222.00      Length: 447
  Ratio: 4.982          Gaps: 0
  Percent Similarity: 99.776 Percent Identity: 96.197

alignment_block:
US-09-303-518D-131 x NQRA_NEIMA ..
Align seg 1/1 to: NQRA_NEIMA from: 1 to: 447
1 ATGATTAATAATCAAAAAGGTCTAAATCTGCCCATCGCGGCGACACCGGA 50
1 MetileysilelylsysGlyLeuasnLeuProillealaglyargProgl 17
51 GCAAGTCATTATGACGGCGCGCCATTACCGAAGTCGCGTTCGTCGGCG 100
17 uGlnValiletyrAspglyProValilleThrGluValAlaLeuLeuGlyG 34
101 AAGNATATCTCGCATCGCCCTCGATGAAATCAAGGAAGGTGAAGCC 150
34 LuGluTyfAlaGlyMeCargproSerMetlylsVallylsGluGlyAspAla 50
151 GTCAAAAAAGCCAAAGTCTGTTTGAAGACAAAAAAGAAATCCGGCGGTAGT 200
51 VallylsysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 ATTACTCGCGCGCTTCAGCAAAATCGCGGTATTACCGTGGCGGCAAA 250
67 lPheThrAlaProValSerGlyLysilleAlaAlaileHisArgGlyGlu 84
251 AGCGGCTACTTCAGTCAGTCGATTCGCGTTGAAGCAACGACGAGAAATC 300
|||||
```

1201 GCGATACCGACGCGCAGGCTTGGTTGGTAAATGGACGAAGA 1250  
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG1 417  
 1251 AGACCTCGCTTGGTGGCAGCTGCTGCTGCGCCGGGCAAAATACGAATACGGCC 1300  
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434  
 1301 CGCTGTTCCGCAAGTCTGCGAACCATTGAGAGGAGGCC 1341  
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447

seq\_name: SwissProt\_40:NQRA\_NEIMB

seq\_documentation\_block:

ID NQRA\_NEIMB STANDARD; PRT; 447 AA.  
 AC Q9K0M3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)  
 DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex  
 DE subunit A) (NQR-1 subunit A).  
 GN NQRA OR NM00569.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Ghinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Clecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RT MC58.";  
 RL Science 287:1809-1815 (2000).  
 CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO  
 CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT  
 CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE  
 CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF  
 CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +  
 CC UBIQUINOL + NA(+) (OUT).  
 CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE  
 CC AND NQRF (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
 CC send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: AE002412; AAF40997.1; -  
 CC TIGR: NM00569; -  
 CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
 CC Complete proteome.  
 CC SEQUENCE 447 AA; 48635 MW; BE174A78439BE477 CRC64;

alignment\_scores:

Quality: 2221.00 Length: 447  
 Ratio: 4.980 Gaps: 0  
 Percent Similarity: 99.776 Percent Identity: 96.197

alignment\_block:

US-09-303-518D-131 x NQRA\_NEIMB ..

Align seg 1/1 to: NQRA\_NEIMB from: 1 to: 447

1 ATGATTAATAATCAAAAAGGTCTTAATCTGCCCATCGCGGCGACAGCGGA 50  
 1 MetileLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProG1 17  
 51 GCAAGTCATTATGACGCGCGCCCATTCACGAAGTCGCGTGTGCGG 100  
 17 uGlnAlaValTyrAspGlyProIleAlaLeuLeuLeuGlyG 34  
 101 AGAATATGTCGCGCATCGCCCTCGATGAAATCAAGGAAGTGAAGCC 150  
 34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50  
 151 GTCAAAAAGGCCAAGTCTGTTTGAAGACAAAAGATCCGGCGGTAGT 200  
 51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67  
 201 ATTACTGCGCGCGCTTCAGGCAAAATCCCGCTATTACCGTGCAGAA 250  
 67 lPheThrAlaProAlaSerGlyLysIleAlaIleHisArgGlyGluL 84  
 251 AGCGCTACTTTCAGTCAGTCGTGATTGCGTTCGAAGGCAACGACGAAATC 300  
 84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluL 100  
 301 GAGTTCGAAAGCTAGCTACCTGACGCGCTGCGGCTGCAAAATTCAGCAGCGA 350  
 101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluG1 117  
 351 AGTGCCTCCCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGACCC 400  
 117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThra 134  
 401 GTCCGTTACGCAAAATCCCTGCGGTAGATGCGGAGCCGCTTCGCCATCTC 450  
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150  
 451 GTCATGCGATGACGACCAATCCGCTGCGTGGCTGCGGACCCCTACGTCATCAT 500  
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIle1 167  
 501 CAAAGAAGCGCGCGAGACTTCAAACGCGGCTGTGTGTTATTGAGCCGCC 550  
 167 elysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184  
 551 TGACCGAAGCTAAATCCATGTGTAAAGCAGCAGCGCGACAGCGTGGCG 600  
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200  
 601 TCTGAAATGCTGCCAATATCAACACATGAATTTGCGGCGCGCATCC 650  
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217  
 651 TCCCGCTTGTAGTGGCAGCGCATTCATTTCATCGAGCCAGTCGCGCGGA 700  
 217 olaglyLeuSerGlyThrHisIleHisPheIleGluProValGlyAla 234  
 701 ATAAACCGTGTGGACCATCAATTAAGACGCTGATTGCTATCGGCGGT 750  
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250  
 751 TTGTTCTGAACGCGCTCTGAATACCGCGCGTGTGCTTGGCTTGGCGG 800  
 251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG1 267  
 801 CTGCAAGTCACAAACCGCGCTCTTCGCTACCGTTCGTTGGTCCGAGG 850  
 267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLys 284

```

851 TGCTCACTACCGCGCGCAATGGTTGACCGGACCAACCGCGTATT 900
284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300
901 TCCGGTTCGGTATTGAACGGTGGGATTCACAGCGCGCCATCATATT 950
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrIle 317
951 GGGACGCTACCAATCAGATTTCCTTATCGAAGAAGCGCGCAGCAAG 1000
317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334
1001 AGCTGTCGGCTGGTGGCGCGCCGCGCAGGACCAATACCTACCGCGC 1050
334 luLeuPheGlyTyrValAlaProGlnProAspLysTyrSerIleThrArg 350
1051 ACCACTCTCGGCATTTCTTAAACAACTCTCAAGTTCACGACAGC 1100
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367
1101 CGTCAACGGCGCGACCGCGCCATCGATCGGCTATGACGCGG 1150
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384
1151 TAATGCGTGGACATCCTCGCTACCTGCTTTGCGCGATTAATCGTC 1200
384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400
1201 GCGCATACCGACAGCGCGCGAGCTTTGGTGTGCTTGAATTTGGACGA 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGlu 417
1251 AGACCTCGCTTTGTCAGCTCTGCTGCAACCATTCAGAAAGGAAGC 1341
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGly 434
1301 CGCTGTTGCGCAAGTCTGCTGCAACCATTCAGAAAGGAAGC 1341
434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447

seq_name: SwissProt_40:NORA_PASMU

seq_documentation_block:
ID NORA_PASMU STANDARD; PRT; 446 AA.
AC Q9CBL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE subunit A) (NQR-1 subunit A).
NORA OR PM328.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PM70;
RC MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -|- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -|- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).

```

-|- SIMILARITY: BELONGS TO THE NORA FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE006171; AAK03412.1; ..  
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
KW Complete proteome.  
SQ SEQUENCE 446 AA; 48168 MW; 5C5287CB77C4932D CRC64;  
  
alignment\_scores:  
Quality: 1700.50 Length: 447  
Ratio: 4.199 Gaps: 1  
Percent Similarity: 90.604 Percent Identity: 73.154  
  
alignment\_block:  
US-09-303-518D-131 x NORA\_PASMU ..  
Align seg 1/1 to: NORA\_PASMU from: 1 to: 446  
  
1 ATGATTAATAATCAAAAAAGGTCTAAATCTGCCCATCGCGGCGACAGCCGA 50  
1 MetIleThrIleLysLysGlyLeuAsnLeuProIleSerGlySerProG 17  
51 GCAAGTCATTTATGACGGCGCGCATTTACCGAAGTCGCGTTCGCTTGGCG 100  
17 uGlnValIleArgAspGlyAsnAlaIleThrGluValAlaLeuLeuGly 34  
101 AGAATATGTCGGCATCGCCCTCGATGAAATCAAGGAAGGTGAAGCC 150  
34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50  
151 GTCAAAAAGGCCAAGTGTCTGTTGAAGACAAAGAAATCCGGCGGTAGT 200  
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67  
201 ATTTACTGGCGCGCTTCAGCAAAATCGCGCTATTCACCGTGGCGAAA 250  
67 lPheThrAlaProAlaSerGlyThrValThrAlaIleHisArgGlyAlaL 84  
251 AGCGCGTACTTCAGTCAGTCGTTGCGGTGAAGCAACGACGAAATC 300  
84 ysArgValLeuGlnSerValValIleLysIleGluGlyAsnGluGlnIle 100  
301 GAGTTCGAACGCTACGTACCTGAACGCGTGGCAAAATTCAGCAGCGAAA 350  
101 ThrPheGluLysTyrThrThrGluGlnLeuAsnGlnLeuThrSerGluG 117  
351 AGTGGCGCCCAACCTGATTCATCAGGCTTATGGACTCGGCTTCGACCC 400  
117 nValArgGlnAsnLeuGlnAlaSerGlyLeuThrAlaLeuArgThr 134  
401 GTCCGTTTCAGCAAAATCCCTCGCGTAGATGCGGCGTTCGCCATCTTC 450  
134 rgProPheSerLysValProAlaValAspAlaThrProValSerIlePhe 150  
451 GTCAATCGCATGACACCAATCCGCTGGCTGCGCCGCTACGGTTCATCAT 500  
151 ValAsnAlaMetAspThrAsnProLeuLysAlaAspProGlnValIleVa 167  
501 CAAGAAGCGCGGAGACTTCAAACGGCGCTGTGGTATTGACGCCGC 550  
167 lGlnGlnSerAlaGlnAlaPheGluAlaGlyLeuThrValLeuSerArgL 184  
551 TGACCCGAACGCTAAATCCATGTGTGTAAAGCAGCAGCGCGCAGCGTCCG 600  
184 euHisGluGlyLysValTyrLeuCysLysAlaAlaAsnAlaSerIlePro 200



alignment\_scores:  
Quality: 1646.00 Length: 448  
Ratio: 4.146 Gaps: 2  
Percent Similarity: 88.616 Percent Identity: 70.089

alignment\_block:  
US-09-303-518D-131 x NORA\_HAEIN ..

Align seg 1/1 to: NORA\_HAEIN from: 1 to: 447

```
1 ATGATTAATAAATAAGGCTAAATCTGCCCATCGCGGACAGCCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 MetilethrilileysylsGlyLeuAspLeuProilealaglylsProal 17
51 GCAAGTCATTATGACGGCCCGCCCATACCGAAGTCGGTTCGTCGGC 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 aGlnValileHisSerGlyAsnAlaValAsnGlnValAlaileLeuGly 34
101 AAGAATATGTCGGCATCGCCCTCGGATGAAATCAAGGAAGGTGAAGCC 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50
151 GTCAAAAAGCCAGTCGTGTTTGAAGACAAAAGAAATCCGGCGTAGT 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValII 67
201 ATTACTGCGCGCTTCAGCAAAATCGCGCTATTCACCGGCGGAAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 ePheThrAlaProAlaSerGlyThrIleThrAlaIleAsnArgGlyGluL 84
251 AGCCGCTACTTCAGTCAGTCGATGTCGCTTGAAGSCAACAGCAATC 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysIle 100
301 GAGTTCGAACGCTACGTACCTCGAAGCCCTGCGAAATTTGACGAGCAA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ThrPheAlaLysTyrSerThrGluGlnLeuAsnThrLeuSerSerGluGI 117
351 AGTCGCGCGCAACTGATTCATCAAGTCAGCTTATGACCTGCGCTTCG 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 nValLysGlnAsnLeuIleGluSerGlyLeuTyrThrAlaLeuArgThr 134
401 GTCGGTTTCAGCAAAATCCCTGCGGTAGATCGCGAGCGTTCGCCATCT 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 rgProPheSerLysValProSerIleGluSerGluAlaSerSerIlePhe 150
451 GTCATCGCATGACACCAATCCGCTCGCTGCGCGACCCCTACGTCATCAT 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValValle 167
501 CAAGAAGACCGCCGGAAGACTTCAACCGCGCTGTTGTTATGAGCGGCC 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 uLysGluTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArgL 184
551 TGACCGAACGCTAA...ATCCATGTGTGTAAGCAGCAGCGCAGAGGTG 597
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 euPheProSerLysProLeuHisLeuLysAlaGlyAspSerAsnIle 200
598 CCGTCTGAAATGTCGCAATATCGAACAACATGAATTTGCGCGCCGCA 647
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ProThrAlaAspLeuGluAsnLeuGlnIleHisAspPheThrGlyValHI 217
648 TCCGTCGGCTGAGTCGACGACATTCATTCATCGACGACGTCGCGC 697
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 sProAlaGlyLeuValGlyThrHisIleHisPheIleAspProValGlyI 234
698 CGAATAAACCCTGGACCATCAATATCAAGACGCTGATTGTTATCGGA 747
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 leGlnLysThrValTrpHisIleAsnTyrGlnAspValIleAlaValGly 250
748 CGTTTGTTCGAACAGCGCTCTGAATACGACGCGCTGTTGCTTGGG 797
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 LysLeuPheThrThrGlyGluLeuTyrSerGluArgValIleSerLeuAl 267
```

```
798 CGGCTCTCAAGTCAACAAACCGCGCTCTTCGCTACCGTTTTTGGTGCGA 847
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 aclyProGlnValLysGluProArgLeuValArgThrThrIleGlyAlaA 284
848 AGGTGCTCAACTTACCGCCGCGGAATTTGGTTGACGCGGACCAACCGGTG 897
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
284 snLeuSerGlnLeuThrGlnAsnGluLeuSerAlaGlyLysAsnArgVal 300
898 ATTTCCGGTTCGGTATTGAACGGTCCGATTGCACAAGCGCGCATGATTA 947
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 ileSerGlySerValLeuCysGlyGlnIleAlaLysAspSerHisAspTy 317
948 TTTGGACGCTACCAACAAATTCAGATTTCGCTTATCGAAGCGCGCAGCA 997
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGluGlyAsnGluL 334
998 AAGACTCTTCGGTGGTTCGGCGGACGCGGACAAATATCTCCATCAGC 1047
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 ysGluPhePheGlyTrpIleMetProGlnAlaAsnLysTyrSerValThr 350
1048 CGCACCACTCTCGGCCATTTCTTAAAAACAACACTCTCAAGTTTCAGC 1097
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 ArgThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThrTh 366
1098 AGCCGCTCAACGCGCGCGCATGTCACCGATCGCGCTTTCGCTTATGAGC 1147
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGluA 383
1148 GCGTAATGCGGTGGACATCTCGCTACCTTCGCTTTCGCGGATTAATC 1197
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
383 rgValMetProLeuAspIleLeuProThrLeuLeuArgAspLeuIle 399
1198 GTCGCGATACCGACGCGCGCGCTTTGGTTCGCTTTCGCTTTCGAGC 1247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
400 ValGlyAspThrAspGlyAlaGlnLeuGlyCysLeuGluLeuAspG 416
1248 AGAAGACCTCGCTTTGTGACGCTTCGCTGCGCGCGGCAAAATACG 1297
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
416 uGluAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrG 433
1298 GCCCGCTTTCGCGCAAGTGTGGAACCATTCAGAAAGGCG 1341
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
433 lysrileLeuArgGlnValLeuAspLysIleGluLysGluGly 447
```

seq\_name: SwissProt\_40:NORA\_VIBHA

seq\_documentation\_block:  
ID NORA\_VIBHA STANDARD; PRT; 446 AA.  
AC Q9RFWJ;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)  
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex subunit A) (NQR-1 subunit A).  
GN NORA.  
OS Vibrio harveyi.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=669;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BB120;  
RX MEDLINE=20056044; PubMed=10587447;  
RA Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,  
RA Gennis R.B., Bogachev A.V., Barquera B.;  
RT "Sequencing and preliminary characterization of the Na<sup>+</sup>-translocating NADH:ubiquinone oxidoreductase from *Vibrio harveyi*."  
RL Biochemistry 38:16246-16252(1999).  
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA<sup>+</sup> IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQR A TO NQR C ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF

CC UBISEMIQUINONE TO UBIQUINOL.  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +  
 CC UBIQUINOL + NA(+)(OUT).  
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE  
 CC AND NQRF (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NQRA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: AFI65980; AAF15411.1; -  
 CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.  
 KW SEQUENCE 446 AA; 48365 MW; 9B05B38BD1D7A97C CRC64;

alignment\_scores:  
 Quality: 1491.50 Length: 447  
 Ratio: 3.854 Gaps: 1  
 Percent Similarity: 86.577 Percent Identity: 64.430

alignment\_block:

US-09-303-518D-131 x NQRA\_VIBHA ..

Align seg 1/1 to: NQRA\_VIBHA from: 1 to: 446

1 ATGATTAAATCAAAAAGGTCTAAATCTGCCCATCGCGGCGACCGGA 50  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1 MetIleThrIleLysLysGlyLeuAlaLeuProIleAlaGlyAlaProSe 17  
 51 GCAAGTCATTATATGACGCGCGCCATTACCGAGTCGCGTTCGTCGGCG 100  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 17 rGlnValIleAsnAspGlyLysThrIleLysLysValAlaLeuLeuGly 34  
 101 AAGAAATATGCGGATCGCCCGCTCGATGAAATCAAGAGAGGTGAAGCC 150  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 34 LuGluTyValGlyMetArgProThrMethHisValArgValGlyAspGlu 50  
 151 GTCAAAAAGCCCAAGTCTGTTTGAAGCAAAAAGAAATCCGGCGTAGT 200  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 51 ValLysAlaGlnValLeuPheGluAspLysLysAsnProGlyVally 67  
 201 ATTACTCGCGCGCTTCAGCAAAATCGCGCTATTACCGCTGGCGAAA 250  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 67 sPheThrAlaProAlaAlaGlyLysValIleGluIleAsnArgGlyAlaL 84  
 251 AGCGGTACTTCACTAGTCAGTCGATGCGGTTGAAGCAACACGAGAAATC 300  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 84 ysArgValLeuGlnSerValIleGluValAlaGlyGluGlnVal 100  
 301 GAGTTCGAACGCTACGTACCTGAGCGCTGGCAAAATTTAGCAGCGAAA 350  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 101 ThrPheAspLysPheGluAlaGlnLeuAlaGlyLeuAspArgGluVa 117  
 351 AGTGGCGCGCAACCTGATTCAATCAGGCTTATGACTCGCGCTTCGACCC 400  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 117 IleLysThrGlnLeuValGluSerGlyLeuTrpThrAlaLeuArgThrA 134  
 401 GTCCGTTCAAGAAATCCCTGCGGTAGATCCGAGCGCGCTTCGCACTTC 450  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 134 rgProPheSerLysValProAlaIleGluSerAlaThrLysAlaIlePhe 150  
 451 GTCAATCGATGACACCAATCGCTGCGCGCGACCCCTACGTCATCAT 500  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 151 ValThrAlaMetAspThrAsnProLeuAlaAlaGlnProGluLeuIleI 167  
 501 CAAAGAAGCCCGCAAGACTTCAAGCGCGCTGTTGGTATTGAGCGCGCC 550  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 167 eThrGluGlnGlnGluAlaPheValAlaGlyLeuAspIleLeuSerAlaL 184

551 TGACCGAAGCTAAATCCATGTGTGTAAGCAGCAGCGCGCAGCGTGGCG 600  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 184 eThrGluGlyLysValTyValCysLys...SerGlyThrSerLeuPro 199  
 601 TCTGAAATGCTGCCAATATCGAAACACATGAATTTGGCGCGCGCATCC 650  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 200 SerSerSerClnSerAsnValGluGluHisValPheAspGlyProHisPr 216  
 651 TGGCGGCTTAGTGGGACGACATTCATTTCATCGAGCGAGTGGCGGGA 700  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValAsnAlaG 233  
 701 ATAAACCGTGGGACCACTATTAAGACGCTGATTGCTATCGGACGT 750  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 233 LuAsnValAlaTrpSerIleAsnTyGlnAspValIleAlaPheGlyGln 249  
 751 TTCTTCGTAACAGCGCTCTGAATACCGAGCGGTGGTTGCTTGGCGGG 800  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 250 LeuPheLeuThrGlyGluLeuTyThrAspArgValValSerLeuAlaG 266  
 801 CTTGCAAGTCAACAAACCGCGCTCTGCGTACCGTTTGGTGGCAAGG 850  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 266 yProValValAsnAsnProArgLeuLeuArgThrValValGlyAlaSerL 283  
 851 TGTCTCAACTTACCGCGCGGGAATTTGGTTGACGCGGACACCGCGTAT 900  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 283 euGluAspLeuThrAspSerGluLeuMetProGlyGluValArgValIle 299  
 901 TCGGTTCTGCTATTGAACGCTGCGATTGCAACAGGCGCGCATGATTAT 950  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 300 SerGlySerValLeuSerGlyThrGlnAlaSerGlyProHisAlaTyrl 316  
 951 GGGACGCTACCAATCAGATTTCGTTATCGAAGAGCGCGCAGCAAG 1000  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 316 uGlyArgTyHisGlnGlnValSerValLeuArgGlyArgAspLysG 333  
 1001 AGCTGTTTCGCGTGGTGGCGCGGACGCGGACAAATCTCCATCATCGCG 1050  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 333 LuLeuPheGlyTrpAlaThrProGlyLysAsnLysPheSerIleThrLys 349  
 1051 ACCACTCTCGGCATTTCTCAAAAACAACTCTTCAAGTTCCACGACAGC 1100  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 350 SerPheLeuGlyHisIlePheLysGlyGlnLeuPheAsnMetThrThr 366  
 1101 CGTCACGCGCGGCGCGCGCATGTTACCGCATCGGCCTTATGACGCG 1150  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 366 rThrAsnGlySerAspArgAlaMetValProIleGlyAsnTyGluArgV 383  
 1151 TAATGCGGTTGGACATCTCTGCTTCTGCTTCTGCGGATTTAATCGTC 1200  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 383 aMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 399  
 1201 GCGGATACGACGCGCGCGCGCTTGGTGGTGGTGGTGGTGGTGGTGG 1250  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 400 GlyAspThrAspSerAlaGlnThrLeuGlyAlaLeuGluLeuAspGlu 416  
 1251 AGACCTCGCTTGTGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1300  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyGlyGlyG 433  
 1301 CGCTGTTGCGCAAGTGTGGAACCATTTGAGAAGGAAGGC 1341  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 433 InLeuLeuArgGluCysLeuAspThrIleValLysGluGly 446  
 seq\_name: SwissProt\_40:NQRA\_VIBAL  
 seq\_documentation\_block:  
 ID NQRA\_VIBAL STANDARD; PRT; 446 AA.  
 AC Q56586;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)



DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)  
 DE (Na(+)-translocating NADH-quinone reductase subunit alpha) (Na(+)-  
 DE translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex  
 DE subunit A) (NQR-1 subunit A).  
 GN NQR OR NQR1.  
 OS Vibrio alginolyticus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=663;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
 RC STRAIN=NCIMB 11036;  
 RA MEDLINE=95104443; PubMed=7805867;  
 RX Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;  
 RT "Cloning and sequencing of four structural genes for the Na(+)-  
 RT translocating NADH-ubiquinone oxidoreductase of Vibrio  
 RT alginolyticus".  
 RL FEBS Lett. 356:333-338(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hayashi M., Unemoto T., Sugiyama A.;  
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-10.  
 RX MEDLINE=98149659; PubMed=9490015;  
 RA Nakayama Y., Hayashi M., Unemoto T.;  
 RT "Identification of six subunits constituting Na<sup>+</sup>-translocating NADH-  
 RT quinone reductase from the marine Vibrio alginolyticus".  
 RL FEBS Lett. 422:240-242(1998).  
 RN [4]  
 RP SEQUENCE OF 1-9 AND 334-340.  
 RX MEDLINE=95104444; PubMed=7805866;  
 RA Hayashi M., Hirai K., Unemoto T.;  
 RT "Cloning of the Na(+)-translocating NADH-quinone reductase gene from  
 RT the marine bacterium Vibrio alginolyticus and the expression of the  
 RT beta-subunit in Escherichia coli".  
 RL FEBS Lett. 356:330-332(1994).  
 RN [5]  
 RP INHIBITION OF ENZYMIC ACTIVITY.  
 RX MEDLINE=20016049; PubMed=10549856;  
 RA Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Unemoto T.;  
 RT "Inhibitor studies of a new antibiotic, korormicin, 2-n-heptyl-4-  
 RT hydroxyquinoline N-oxide and Ag<sup>+</sup> toward the Na<sup>+</sup>-translocating NADH-  
 RT quinone reductase from the marine Vibrio alginolyticus".  
 RL Biol. Pharm. Bull. 22:1064-1067(1999).  
 RN [6]  
 RP REVIEW.  
 RX MEDLINE=21145117; PubMed=11248187;  
 RA Hayashi M., Nakayama Y., Unemoto T.;  
 RT "Recent progress in the Na(+)-translocating NADH-quinone reductase  
 RT from the marine Vibrio alginolyticus".  
 RL Biochim. Biophys. Acta 1505:37-44(2001).  
 RN [7]  
 RP REVIEW.  
 RX MEDLINE=21145118; PubMed=11248188;  
 RA Steuber J.;  
 RT "Na(+)-translocation by bacterial NADH:quinone oxidoreductases: an  
 RT extension to the complex-I family of primary redox pumps".  
 RL Biochim. Biophys. Acta 1505:45-56(2001).  
 CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO  
 CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT  
 CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQR TO NQRE  
 CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF  
 CC UBISEMIOQUINONE TO UBIQUINOL.  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +  
 CC UBIQUINOL + NA(+) (OUT).  
 CC -1- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+) +  
 CC PUMPING ACTIVITY AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.  
 CC INHIBITED BY KORORMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE  
 CC (HONO).  
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRE  
 CC AND NQRF.  
 CC -1- SIMILARITY: BELONGS TO THE NQRA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; Z37111; CAA85476.1; -.  
 DR EMBL; AB008030; BAA22910.1; -.  
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.  
 FT CONFLICT 337 337 W->L (IN REF. 4).  
 SQ SEQUENCE 446 AA; 48622 MW; 6D69ACAA53FE515C CRC64;  
 CC -----  
 alignment\_scores:  
 Quality: 1488.50 Length: 447  
 Ratio: 3.856 Gaps: 1  
 Percent Similarity: 86.353 Percent Identity: 63.758  
 alignment\_block:  
 US-09-303-518D-131 x NQRA\_VIBAL ..  
 Align seg 1/1 to: NQRA\_VIBAL from: 1 to: 446  
 1 ATGATTAAATCAAAAAAGGTCTAAATCTGCCATCTCGCGGCGAGACCGGA 50  
 1 MetIleThrIleLysLysGlyLeuAspLeuProLleAlaGlyThrProSe 17  
 51 GCAAGTCATTATGACGCGCGCCGTCATTACGAGTCGCGTCTGCTGGCG 100  
 17 rGlnValIleAsnAspGlyLysLysValAlaLeuLeuLeuGly 34  
 101 AAGAATATCTCGGCATGCGCCCTCGATGAAATCAAGGAAGGTGAAGCC 150  
 34 lueGluTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50  
 151 GTCAAAAAAGCCAAAGTGTCTTTGAAGACAAAAAGATCCGGCGGTAGT 200  
 51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyVal 67  
 201 ATTACTGCGCGGTCTCAGGCAAAATCGCCGCTATTACCGTGGCGAAA 250  
 67 sPheThrAlaProAlaAlaGlyLysValIleGluValAsnArgGlyAlaL 84  
 251 AGCGCGTACTTCAGTCAGTCGCTGATTGCGTTGAAGCAACGACGAAATC 300  
 84 ysArgValLeuGlnSerValValIleGluValAlaGlyGluGlnVal 100  
 301 GAGTTCGAACGCTACCTACCTGAGCGGTGGCAAAATTCAGCAGCGAAA 350  
 101 ThrPheAspLysPheGluAlaAlaGlnLeuSerGlyLeuAspArgGluVa 117  
 351 AGTGGCGCCCACTGATTCAATCAGGCTTATGAGTGGCTTCGACCC 400  
 117 lIleLysThrGlnLeuValAspSerGlyLeuThrAlaLeuAlaGlyThr 134  
 401 GTCGCTTCAGCAAAATCCCTCGCTAGATGCGGAGCGGTCGCCATCTTC 450  
 134 rgProPheSerLysValProAlaIleGluSerSerThrLysAlaIlePhe 150  
 451 GTCAATGCGATGACACCAATCGCTGGCTGCGGCGCCCTACGTCATCAT 500  
 151 ValThrAlaMetAspThrAsnProLeuAlaAlaLysProGluLeuIle 167  
 501 CAAAGAGCGCGCAAGACTTCAAAGCGCGCTGTTGGTATTGAGCGCGC 550  
 167 easnGluGlnGlnGluAlaPheIleAlaGlyLeuAspIleLeuSerAlaL 184  
 551 TGACCGAAGCTAAATCCATGTGTAAAGCAGCAGCGCGCAGCGTGGCG 600  
 184 euThrGluGlyLysValTyrValCysLys...SerGlyThrSerLeuPro 199



```

601 TCTGAAATGCTGCAATATGAAACACATGAATTTTGGCGCGCATCC 650
      ::::: ::::: ||| ||| ||| ||| ||| ||| |||
200 ArgSerGlnSerAsnValGluGluHisValPheAspGlyProHispr 216
      ::::: ::::: ||| ||| ||| ||| ||| ||| |||
651 TCCGGCTGAGTGGCGCGCACATTCATTCATCGAGCCAGTCGCGCGA 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValAsnAlaG 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 ATAAACCGTGTGACCATCAATATCAAGACGTCATTCGTCAGCGT 750
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
233 luAsnValAlaTrpSerIleAsnTrpGlnAspValIleAlaPheGlyLys 249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
751 TTGTTCTGTAACAGCGCTCTGAATACGAGCGCGTGTGCTTGGCGG 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 LeuPheLeuThrGlyGluLeuTyThrAspArgValValSerLeuAlaG1 266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 CCTGCAAGTCAACAACCGCGCTCTTTCGCTACCGTGGTGGTGGCGA 850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 yProValValAsnAsnProArgLeuValArgThrValIleGlyAlaSerL 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
851 TGTCTCAACTTACCGCGCGGAATTTGTTGACGCGGACACCGCGTATT 900
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
901 TCCGGTTCGGTATTGAACGGTTCGATTGCACAGCGCGCATGATTATT 950
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 SerGlySerValLeuThrGlyThrHisAlaThrGlyProHisAlaTyLe 316
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 GGGACGTACACATCAGATTCCTCGTTATCGAAGAGCGCGGCAAG 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 uGlyArgTrpHisGlnValSerValLeuArgGluGlyArgGluLysG 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 AGCTGTTTCGGTGGTGGCGCGAGCGCGCAAAATCTCATCACGCGC 1050
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 ACCACTTCGGCATTTCTTAAACAACTCTTCAAGTTCAGCAGC 1100
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 SerPheLeuGlyHisValPheLysGlyGlnLeuPheAsnMetThrThr 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 CGTCAAGCGGGCGGACCGCGCATGTGTACCGTGGCCTATGAGCGCG 1150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTrpGluArgV- 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1151 TAATGCGTGGACATCCTCGTACCTTGTGTTGCGCGATTAATCGTC 1200
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1201 GGCATACCGACGCGCGAGCTTGGTGTGCTTGGATGGACGAAGA 1250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 GlyAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGluLeuAspGluG1 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1251 AGACCTCGCTTTGTGACGCTCTGCTGCCCGGCAATACGAATACGGCC 1300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyThrGlyTyGlyT 433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1301 CGCTGTGCCCAAGTCTGGAAACCATTCGAGGAAGGAGCC 1341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 hrLeuLeuArgGluCysLeuAspThrIleGluLysGluGly 446
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq\_name: Swissprot\_40:NQRA\_VIBCH

seq\_documentation\_block:  
 ID NQRA\_VIBCH STANDARD; PRT; 446 AA.  
 AC Q9KPS1; Q9X4Q3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)  
 DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A). (NQR complex  
 DE subunit A) (NQR-1 subunit A).  
 DE NQRA OR VC2295.  
 OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=99179036; PubMed=10077658;  
 RA Haese C.C., Mekalanos J.J.;  
 RT "Effects of changes in membrane sodium flux on virulence gene  
 expression in Vibrio cholerae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3183-3187(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO  
 UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT  
 OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE  
 ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF  
 UBISEMIOQUINONE TO UBIQUINOL.  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +  
 UBIQUINOL + NA(+) (OUT).  
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE  
 AND NQRF (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NQRA FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF117331; AAD29962.1;  
 DR EMBL; AE004300; AAF95439.1; ALT\_INIT.  
 DR TIGR; VC2295;  
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
 KW Complete proteome.  
 SQ SEQUENCE 446 AA; 48624 MW; 428E8C397EBA163D CRC64;

alignment\_scores:

Quality: 1476.50 Length: 447  
 Ratio: 3.845 Gaps: 1  
 Percent Similarity: 85.906 Percent Identity: 62.416

alignment\_block:

US-09-303-518D-131 x NQRA\_VIBCH ..

Align seg 1/1 to: NQRA\_VIBCH from: 1 to: 446

```

1 ATGATTAATAAATCAAAAAGGTCATAATCTGCCATCGCGGCGACACGGA 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 GCAAGTCATTATGACGGCGCGGCATTCACGAAGTCGCGTGTGCTGCGC 100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGlyG 34
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 AAGAATATGTGCGCATGCGCCCTCGATGAGAAATCAAGGAAGGTGAAGCC 150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 luGluTyrvValGlyMetArgProThrMetHisValArgValGlyAspGlu 50

```

151 GTCAAAAGGCGCAAGTGTGTTTGAAGACAAAGAAATCCGGCGCTACT 200  
|||||  
51 Valysylalaglnlleleuphegluaspllyslasproglyvally 67  
201 ATTACTGCGCGCGTTCAGGCAAAATCCCGCTATTACCGCTGGCGAAA 250  
|||||  
67 sPheThrSerProValserGlyLysValValGluilleAsnArgGlyAlaL 84  
251 AGCGGTACTTACGTAGTCGTGATTCGCGCTTGAAGCAAGCAAGCAATC 300  
|||||  
84 ysargValLeuGlnSerValValIleGluValAlaGlyAspGlnVal 400  
301 GAGTGTGAAGCGTACGTACCTGAGCGCTGGCAAAATTTGAGCAGCAAAA 350  
|||||  
101 ThrPheAspLysPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 117  
351 AGTGGCGCGCAACCTGATTCATCAATCAGCTTATGACTGGCTTTCGACCC 400  
|||||  
117 aileLysThrGlnLeuValGluSerGlyLeuTrpThrAlaPheArgThra 134  
401 GTCCGTTACGAAATCCCTGCGGTAGATGCCGAGCGCTGCGCATCTTC 450  
|||||  
134 rgProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 150  
451 GTCAATCGCATGGACACCAATCCGCTGGCTGCGGACCTACGCTCATCAT 500  
|||||  
151 ValThrAlaMetAspThrAsnProLeuAlaAlaGluProThrValVal 167  
501 CAAAGAGCGCGGAGACATTCACACCGCGCTGTGTGTTATTGAGCGGCC 550  
|||||  
167 eAsnGluGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAlaL 184  
551 TGACCGAAGCTTAAATCCATGTGTGTAAACGACGCGCGCAGCGTGGCG 600  
|||||  
184 eutThrGlyLysValValValCysLys...LysGlyThrSerLeuPro 199  
601 TCTGAAATGTCGCAATATCGAACACATGAATTTGGCGCGCGCATCC 650  
|||||  
200 ArgSerGlnGlnProAsnValGluGluHisValPheAspGlyProHisPr 216  
651 TGCGCGCTGAGTGGCGGACCATTCATTCATCCGACCGCAGTCGCGCGA 700  
|||||  
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValSerAlaA 233  
701 ATAAACCGTGTGACCATCAATATCAACGCGCTGATTCATCGGACGT 750  
|||||  
233 sPHisValAlaIleTrpSerIleAsnTyroGlnAspValIleAlaValGlyGln 249  
751 TTGTCGTAAACGCGCTCTGTAATACCGAGCGGTGTTGCTTGGTGGCGG 800  
|||||  
250 LeuPheLeuThrGlyGluLeuTyThrGlnArgValValSerLeuAlaG1 266  
801 CCTGCAAGTCAACAAACGCGCTTTCGCTACCGCTTTGGTGGCGAAGG 850  
|||||  
266 yProValValAsnLysProArgLeuValArgThrValMetGlyAlaSerL 283  
851 TGCTCACTTACCGCGCGGAATTTGTTGACGCGGCAACCGCGTGAAT 900  
|||||  
283 euGluGlnLeuValAspSerGluileMetProGlyGluValArgIlele 299  
901 TCCGTTGCGGTATTGACGGTGCATTCACACGCGCGCATGATTATT 950  
|||||  
300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyLe 316  
951 GGGAGCTTACCACATCAGATTTCCGTTATCGAAGAGCGCGCAGCAAG 1000  
|||||  
316 uGlyArgTyThrHisLeuGlnValSerValLeuArgGluGlyArgAspLysG 333  
1001 AGCTTTCGCTGGGTTCGCGCGGCGGCAAAATATCATCATCAGCGCG 1050  
|||||  
333 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349  
1051 ACCACTCTCGGCCATTTCTTAAAAACAACTCTTCAAGTTCACGACAGC 1100

350 SerPheLeuGlyHisLeuPheLysGlyGlnValTyAsnMetThrThr 366  
|||||  
1101 CGTCAACGCGCGGACCGCATCGTACCGATCGGACCTTATGAGCGCG 1150  
|||||  
366 rThraSnglySerAspSerMetValProIleGlyAsnTyroGlyLysV 383  
1151 TAATGCCCTTGGACATCTCTACCTTGTCTTTCGCGGATTTAATCGTC 1200  
|||||  
383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLysAla 399  
1201 GCGCATACGACGCGCGCAGCGCTTGGTTCCTTGAATTTGGACGAGA 1250  
|||||  
400 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGluLeuAspGluG1 416  
1251 AGACCTCGCTTGTGACGCTCGTCCGCGGCAAAATACGAATACGCGCC 1300  
|||||  
416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyroGlyGlyG 433  
1301 CGCTGTCGCAAAAGTGTGGAACCATTCATGAGAGGAGGC 1341  
|||||  
433 InLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 446  
seq\_name: SwissProt\_40:NORA\_PSEAE  
seq\_documentation\_block:  
ID NORA\_PSEAE STANDARD; PRT: 445 AA.  
AC Q9HZK6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Na(+)-translocating NADH-quinone reductase subunit A (BC 1.6.5.-)  
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex  
DE subunit A) (NQR-1 subunit A).  
GN NORA OR PA2999.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover K.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
"Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO  
OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE  
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF  
CC UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) -> NAD(+) +  
CC UBIQUINOL + NA(+)(OUT).  
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRE, NQRC, NQRD, NQRE  
CC AND NQRF (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE004724; AAG06387.1;  
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
KW

KW Complete proteome.  
SQ SEQUENCE 445 AA; 48081 MW; ED0468D0659B701 CRC64;

alignment\_scores:  
Quality: 1358.00 Length: 447  
Ratio: 3.641 Gaps: 1  
Percent Similarity: 83.445 Percent Identity: 59.060

alignment\_block:

US-09-303-518D-131 x NORA\_PSEAE

Align seg 1/1 to: NORA\_PSEAE from: 1 to: 445

```
1 ATGATTAAAAATCAAAAGGCTAAATCTCCCATCGCGGCGAGACGGGA 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetilelysileysargGlyLeuAspLeuProIleSerGlyAlaProGI 17
51 GCAAGTCATTATGACGGCCCGCCATACCGAAGTCGGCTTGGCG 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGlnArgIleGluAlaAlaArgProValArgSerValAlaLeuIleGlyP 34
101 AAGAATATGTCGCATCGCCCTCGATGATAAATCAAGGAAGTGAAGCC 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 heAspTyrHisGlyMetLysProThrMetAlaValGlnValGlyAspArg 50
151 GTCAAAAAGGCGCAAGTCGTGTTGAAGACAAAGAAATCCGGCGTAGT 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValLysLeuGlyGlnValLeuPheThrAspLysLysAsnProSerValse 67
201 ATTACTGCGCGCGCTCAGGCAAAATCGCGCTATTCACCGTGGCGAA 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 rTyThrAlaProGlyValGlyValSerAlaIleHisArgGlyGluL 84
251 AGCGCGTACTTTCAGTCAGTCGTGATTCGCTTGAAGGCAAGCAAGAAATC 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysArgValLeuGlnSerValValIleAspLeuAspGlyAspGlnGlnLeu 100
301 GAGTTGACAGCTACTACTGAGCGCTGAGCGCTGGCAAAATGAGCAGCAAA 350
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 GluPheAlaArgTyrProAlaAspLysLeuAlaThrLeuSerAlaGluGI 117
351 AGTGGCGCGCAACTGATTCATCACTGAGCTATGAGCTGCGCTTCGCAACC 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValArgAspAsnLeuGlnSerGlyLeuThrAlaLeuArgThrA 134
401 GTCGCTTCACAAAAATCCCTGCGTAGATGCGGAGCGCTTCGCATCTTC 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rgProPheSerLysValProAspProGluSerSerProSerIlePhe 150
451 GTCATGCGATGACACCAATCCGCTGCGTGGCTGCGACCCCTACGCTCATCAT 500
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValThrAlaIleAspThrGlnProLeuAlaAlaAspProGlnValValII 167
501 CAAGAAGCGCGGAGAGACTTCAACCGCGCGCTGTTGGTATTGACGCGCC 550
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eAlaGluGlnGlyGluAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184
551 TGACCCAGACGTAATATCCATGTGTGTAAAGCAGCAGCGCACGCTGGCG 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euAla.....ArgValPheLeuCysLysAlaGluGlyValSerLeuPro 198
601 TGTGAATATGTCGAATATCAATATCAATATTCATTCATTCAGCGCGCATCC 650
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GlyGluAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisPr 215
651 TGCGCGCTTGTGTCAGCAGCATTCATTCATTCATTCAGCGCGCGGGA 700
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlaG 232
701 ATAAACCGTGTGGACCATCAATATCAAGACGTGATTCGCTATCGGAGCT 750
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 lYlYSerValTrpAsnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248
```

```
751 TTGTTGTAACAGCGCTGTAATACCGAGCGCGTGGTTCCTTGGCGCG 800
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 LeuPheThrThrGlyGlnLeuTrpThrGluArgValIleAlaLeuAlaGI 265
801 CCTGCAAGTCAACAAACGCGCTCTCGTACCGTTTTTGGTGCGAAGG 850
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 yProValValGluLysProArgValValArgThrArgLeuGlyAlaAsnL 282
851 TGTCTCAACTTACCGCGCGGAATTTGGTGTGCGGACGACCGGTGATT 900
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 euAspGluLeuAlaAlaGlyGlnLeuGlnProGlyAsnAsnArgLeuIle 298
901 TCCGTTTCGGTATTGAACGGTCGATTTCCTGTTTGAAGAAGCGCGCAAG 1000
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 SerGlySerValLeuGlyArgThrAlaHisGlyAlaTyrAlaTyrLe 315
951 GGGAGCTACCAACATCAGATTTCCTTAAACAACTCTTCAAGTTTCACGAC 1000
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 uGlyArgTyrHisLeuGlnLeuSerCysLeuLysGluGlyAspGlnArg 332
1001 ACCTGTTTCGGTGGTTCGCGCGAGCGGACAAATACCTCATCATCGCGC 1050
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 luPheLeuHisTyrLeuArgAlaGlyValGluLysHisSerLeuLeuAsn 348
1051 ACCACTCTCGGCGCATTTCTTAAACAACTCTTCAAGTTTCACGACAGC 1100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 ValPheValSerArgLeuLeuGlyLysArgPheAlaPheThrThrSe 365
1101 CCGTCAACGCGCGCGCATGATGATGATGATGATGATGATGATGATGATG 1150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 rThrAsnGlySerProArgAlaMetValProValGlyAsnTyrGluAlaV 382
1151 TAATCCCTTGGACATCTGCTACCTACCTACCTACCTACCTACCTACCT 1200
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 alMetProLeuAspIleLeuProThrGlnLeuLeuArgTyrLeuIleVal 398
1201 GCGCATACGACAGCGCGCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGG 1250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 GlyAspThrGluMetAlaGlnLysLeuGlyAlaLeuGluLeuAspGluGI 415
1251 AGACCTCGCTTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 uAspLeuAlaLeuCysSerTyrValCysAlaGlyLysTyrGluTyrGlyP 432
1301 CGCTGTTGCGCAAGTCTGGAACCATTTGAGAGGAGGAGG 1341
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 rolLeuArgAspAsnLeuAlaArgIleGluGlnGluGly 445
seq_name: SwissProt_40:NORA_CHLMD
seq_documentation_block:
ID NORA_CHLMD STANDARD; PRT; 465 AA.
AC Q9PLD3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable Na(+)-translocating NADH-quinone reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE (NQR complex subunit A) (NQR-1 subunit A).
GN NORA OR TC0002.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman R., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
```





700 AATAAA...ACCGTGTGACCAATCAATTAACAAGACGTGATGCTATCGG 746  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
240 GluLysGluValAlaPheThrLeuSerPheGlnAspValLeuThrIleG1 256  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
747 ACGTTTGTCTGAACAGCGCTGTAATACGAGCGGTGTTGCTGCTGG 796  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
256 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 273  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
797 CGGCG...CTCAAGTCAACAACCGCGCTCTGCGTACCGTTTG 840  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
273 laGlyThrAlaLeuLysSerLeuArgArgTyrValIleThrLys 289  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
841 GTGCGAAGTGTCTCAACTACCGCGGGAATTGTTGACGCGGACAA 890  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
290 GlyAlaSerPheSerLeuIleAsnLeuAsnAspIleSerAsnAs 306  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
891 CGCGGTGATTCGGTGGTATTGAACGGTGCATTGACAAAGCGCGC 940  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
306 pThrLeuIleSerGlyAspProLeuThrGlyArgLeuCysLysGluG 323  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
941 ATGAT...TATTGGAGCGCTACCAATCAGATTTCGGTTATCGAAGA 987  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
323 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 339  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
988 GCGCGCAGCAAGAGCTGTCGGTGGTTCGCGCGCAGCGGACAAATA 1037  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
340 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 356  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1038 CTCATACAGCGCGACCACTTCGCGCATTCCTTAAACAACTCTCA 1087  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
356 oThrPheThrLysThrTyrLeuSerGlyPhePheLysLysArg...T 372  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1088 AGTTCACG...ACAGCGCTCAACGCGCGCGCGCGCATGGTA 1128  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
372 hrTyrThrAsnProAspThrAsnLeuHisGlyLeuThrArgProIleIle 388  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1129 CGATCGGCACCTATGACGCGGTAAATCGCTGGACATCGCTACCT 1178  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
389 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 405  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1179 GCTTTTGGCGGATTAATCGCTCGCGATACCGACGCGCGCGCTTGG 1228  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
405 oleuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 422  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1229 GTTGTCTGGAATTGACGAAGAAGACCTCGCTTTGTCAGCTTCGTC 1278  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
422 lyPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleasp 438  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1279 CGGCGCAATACGAATACGCGCGCTGTGCGCAAGTGTGGAACCAT 1328  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
439 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGlu 455  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1329 TCAGAAGGAA 1338  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
455 rAlaLysGlu 458  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

seq\_name: SwissProt\_40:NQRA\_CHLTR

seq\_documentation\_block:

ID NQRA\_CHLTR STANDARD; PRT; 465 AA.

AC 084639;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable Na(+)-translocating NADH-quinone reductase subunit A

DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)

DE (NQR complex subunit A) (NQR-1 subunit A).

GN NQRA OR CT634.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=813;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D/UW-3/CX;

RX MEDLINE=99000809; PubMed=9784136;

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:

RT Chlamydia trachomatis.";

RL Science 282:754-759(1998).

CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO

CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT

CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE

CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF

CC UBISEMINONE TO UBIQUINOL (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +

CC UBIQUINOL + NA(+) (OUT).

CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE

CC AND NQRF (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE NQRA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE001334; AAC68238.1; -

KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;

KW Complete proteome.

SQ SEQUENCE 465 AA; 51757 MW; 303F88DC133AB998 CRC64;

alignment\_scores:

Quality: 432.00 Length: 464

Ratio: 1.474 Gaps: 14

Percent Similarity: 63.147 Percent Identity: 28.017

alignment\_block:

US-09-303-518D-131 x NQRA\_CHLTR ..

Align seg 1/1 to: NQRA\_CHLTR from: 1 to: 465

4 ATTAATAACAAAAGTCTAATCTGCCATCGCGGCGGACGACGAGCA 53

||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

3 IleValValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysG1 19

: :

54 AGTCATTATGACGCGC.....CCGGCCATTACCGAAGTCGCGTTCG 94

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

19 uSerGlyPheCysGlyLysValAspProThrTyrValSerValAspLeu. 35

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

95 TTGCGCAAGAATAATATGTCGCGCATCGCCCTCGATGAAATCAAGGAAGT 144

: :

36 .....ArgProPheAlaProLeuProLeuGlyValLysValThrProGlu 50

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

145 GAAGCGCGTCAAAAAGCGCAAGTCGCTTTGAAGACAAAGAAATCCGGG 194

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

51 AspGlnValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerG1 67

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

195 CGTAGTATTACTGCGCGCTTCAGGCAAAATCGCGCTATTACCGCTG 244

||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

67 yValPheIleThrSerProValAspGlyGluValValGluIleArgArg 84

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

245 GCGAAAAGCGGTACTTCAGTCAGTCGCTGATTCGCCGTTGAA...GCCAAC 291

||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

84 lysAsnLysArgAlaLeuLeuGluIleValIleLysLysProGlyIle 100

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

292 GACCAATCGAGTTCGAACGCTACGTACCTGGAAGCGCTGCGCAAAATGAG 341

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

101 SerGlnThrLysPheSer.....TyrAspLeuGlnSerLeuTh 113

: : : : : ||| : : : : : ||| : : : : : ||| : : : : :

342 CAGGAAAAAGTGGCGCGCAACCTGATTCAATCAGGCTTATGACTGCGC 391

412 uThrGlnAsnPhcGlUGluAlaCysArgLeuGlyLeuLeuGluValalAp 429

1247 AAGAAGACACCTCGCTTTGTGTCAGCTTCCTCTGCCCGGCAATAACGAAATC 1296

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

429 roGluAspPheAlaLeuProThrPheIleaspProSerLysThrGluMet 445

1297 GGGCCCTCTTGGCGAAAGTCTGTGGAACACCATTTGAGAAGAA 1338

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

446 PheSerIleValIlyGluSerLeuLeuArgTyrAlaLysGlu 459

seq\_name: SwissProt\_40:RNFC\_VIBCH

seq\_documentation\_block:

ID	RNFC_VIBCH	STANDARD;	PRT;	774 AA.
AC	QK188;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Electron transport complex protein infc.			
GN	RNFC OR VC1015.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;			
ON	NCBI_TaxID=666;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;			
RX	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayson			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson			
RA	Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I.,			
RA	McDonald L., Uteback T., Fleischmann R.D., Nierman W.C., V			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., V			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen			
RT	cholerae.";			
RL	Nature 406:477-483(2000).			
CC	-1- FUNCTION: May be part of a membrane complex involved in			
CC	transport (By similarity).			
CC	-1- COFACTOR: Binds 2 4FE-4S clusters (Potential).			
CC	-1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB,			
CC	rnfC, rnfE and rnfG (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Inner membrane-associated (By simi			
CC	-1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOX			
CC	RNFC SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a			
CC	between the Swiss Institute of Bioinformatics and the EMB			
CC	the European Bioinformatics Institute. There are no restr			
CC	use by non-profit institutions as long as its content is			
CC	modified and this statement is not removed. Usage by and			
CC	entities requires a license agreement (See <a href="http://www.isb-s">http://www.isb-s</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; AE004183; AAP94176.1; ALT_INIT.			
CC	TIGR; VC1015; -			
DR	InterPro; IPR001450; 4Fe4S_ferredoxin.			
DR	InterPro; IPR001949; Complex1_51K.			
DR	Pfam; PF001512; Complex1_51K; 1.			
DR	Pfam; PF00037; fer4; 2.			
DR	PROSITE; PS00198; 4FE4S-FERREDOXIN; 2.			
KW	Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;			
KW	Complete proteome.			
FT	METAL 378 378			
FT	METAL 381 381			
FT	METAL 384 384			
FT	METAL 388 388			
FT	METAL 417 417			
FT	METAL 420 420			
FT	METAL 423 423			
FT	METAL 427 427			
FT	SEQUENCE 774 AA; 83500 MW; 44EB13EC304E334AE CRC64;			
SQ				



alignment\_scores:  
Quality: 172.00 Length: 472  
Ratio: 0.782 Gaps: 17  
Percent Similarity: 46.610 Percent Identity: 21.398

alignment\_block:  
US-09-303-518D-131 x RNFC\_VIBCH ..  
Align seg 1/1 to: RNFC\_VIBCH from: 1 to: 774

```
52 CAAGTCAATTATGACGCGCGCGCATACCGAAGTCGCGTTCGTCGCGA 101
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32 GlnProIleIleAsnAlaSerIleProAsnGluLeuValLeuProLeuLeu 48
102 AGAATATGTCGGATCGCGCCCTCGATGATGAATAACAGGAAGTGAACCG 151
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 sGlnHisIleGlyAlaGlyAspLeuLeuValValValGlyAspArgV 65
152 TCAAAAAGCGCAAGTCTGCTTTGAAGACAAAAGAAATCGGCGGTAGTA 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 allLeuLysGlyGlnProLeuThrGlnIleThrSerThrPheMetLeuPro 81
202 TTTACTGCGCGCGTTCAGCAAAATCGCGCTATTACCGTGGCGAAAAA 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 IleHisAlaProThrSerGlyValIleSerAlaIle.....GluPr 95
252 CGCGTACTTCAG.....TCAGTCTGCTGATTG 277
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 oArgThrValAlaHisProSerGlyLeuSerGluLeuCysIleValLeu 112
278 CCGTTGAAGCAAGACGAA.....ATCGAGTTC 306
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 hrProAspGlnGlnGluTrpPheGluLeuGlnProGlnProAspPhe 128
307 GAACGCTACGTACTGAAGCGTGGCAAAATTCAGCAGCAAAAGTGGC 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129 GlnGlnLeuThrProGluThrLeuLeuGluLeuIleArgGlnAlaGlyII 145
357 CCGCAACTGATCAATCAGCGTATGAGTTCGCGTTCGCGACCGCCGT 406
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 eSerGlyMetGlyGlyAlaGlyPheProThrAla..... 156
407 TCAGCAAAATCCCTGCGGTAGATCGCGAGCGTTCGCGCATCTCGTCAAT 456
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 ..LysLysLeuGlnSerGlyLeuSerArgThrGluIleLeuIleAsn 172
457 GCGATGAGACACCAATCCGCTGGTGGCGCGACCCCTACGGTCAATCAAGA 506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 AlaAlaGluCysGluProTyrIleThrAlaAspValLeuMetArgG1 189
507 AGCGCGCAAGACTTCAACGCGCGCTTGGTATTGAGCGCGCTGACCG 556
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 nTyrAlaHisGluIleGlnGlyIleGluIleValGluHisIleLeuL 206
557 AACGTAATAATCCATGTG..... 573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 ysProLysLeuThrIleIleGlyLeuAspAsnLysProGluAlaVal 222
574 .....TGTAAGCAGCAGCGCGCAGACGTGCGGTCT..... 603
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 AlaAlaLeuGlnAlaAlaGlnAspLysProMetValIleArgValII 239
604 .....GAAATCCTG 613
239 eProThrLysTyrProSerGlyGlyGlyLysGlnLeuIleIleLeuT 256
614 CCAATATGAAACACATGAATTTGGCGCGCGCATCTCGCGCTGAGT 663
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 hrAsnLeuGluValProLysGlyIleProAlaAspIleGlyLeu... 271
664 GGCAGGCACATTCATTCATCGACCGAGTGGCGCGGAATAAACCGGTG 713
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
272 .....MetValGlnAsnValGlySer..... 278
714 GACCATCAATTATCAAGACGTGATTGTCATCGGACGTTTGTTCGTAACAG 763
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
279 .....LeuGlnAlaIleAlaArgAlaIleValHisG 289
764 GCGCTCTGAATACGAGCGCGTGGTTCCTTGGCGCGCTGCAAGTCAAC 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 lyGluProLeuIleArgValValThrLeuThrGlyAspCysPheArg 305
814 AAACCGCGCTCTTCGCTACCGTTTGGTTCGAGGTGCTCAACTTAC 863
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 LysProArgAsnValTrpAlaLeuLeuGlyThrProValGlnAlaLeuLe 322
864 CGCGCGCGAATTGGTTGACGCGGACAA.....CGCGTGAATTCG 904
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 uAsnGluPheGlyTyrLysAlaAspLysLysLeuProArgLeuIleMetG 339
905 GTCGCTGATTGAACGGTTCGATTGCACAAAGCGCGCATGATTATTGGGA 954
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 lyGlyProMetMetGlyPheThrLeuPro..... 348
955 CGCTACCAACATCAGATTTCGTTATCAAGAGCGCGCAGCAAGAGCT 1004
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 .....HisAlaGlnValProIle.....ThrLysThrAl 358
1005 GTTCGGTGGTTCGCGCGCAGCGGCAAAATACTCCATCAGCGCGCACCA 1054
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 aAsnCysIleLeuAlaPro.....ThrArgAsnG 368
1055 CTCCTCGGCCATTCTCTAAACAAACAACTCTCAAGTTCACGACAGCCGTC 1104
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 luLeu.....ThrSerSerAsp 373
1105 AACGCGCGCGACCGCCATGTCGCGCATCTATGAGCGCGTAAT 1154
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 Asn.....GluMetalcysIleArgCysGlyGlnCysAlaGluAlaC 388
1155 GCGCTTGACATCTCCCTGCTTCCTTTTGGCGGATTAATCGTCGCGC 1204
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 sProValSerLeuLeuProGlnGlnLeuGlnTrpHisAlaLysAlaGlu 405
1205 ATACGACAGCGCGCAGGCTTGGGTTCGTTGGAAATTCGAGCAAGACAG 1254
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 luPheAspLys.....CysGluGluLeuAspLeuLysAsp 416
1255 .....CTGCTTGTGCAAGTTCGCTGCGCGGCAAAATACGAATA 1295
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 CysIleGluCysGlyAlaCysAlaIleValCysProSerGluIle..... 431
1296 CGGCGCGCTGTTGCGC 1311
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 .....ProLeuValGln 435
seq_name: SwissProt_40:YM96_YEAST
seq_documentation_block:
ID YM96_YEAST STANDARD; PRT; 1140 AA.
AC Q04893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN YMR317W OR YM924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RL
```



alignment_scores:	Quality: 135.50	Length: 465
	Ratio: 0.622	Gaps: 15
Percent Similarity:	46.862	Percent Identity: 19.570
alignment_block:	US-09-303-518D-131 x RNFC_HAEIN	..
Align seg 1/1	to: RNFC_HAEIN	from: 1 to: 819



```
142 laGlyPheProThrAlaAlaLysLeuAlaAlaArgProAlaGluLysIle 158
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
418 CTGCGGTAGATGCGGAGCGGTTGCGCACTTCGTCATGCGATGCGACAC 467
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
159 .....HisThrLeuValValAsnGlyAlaGluCy 168
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
468 CAATCGGTGGTGGCGGACCCCTACGGTCATCATCAAGAACGCGCGAAG 517
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
168 sGluProTyrIleSerAlaAspLeuLeuMetArgGluAlaThrG 185
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
518 ACTTCAACGCGCGCTGTTGGTATTGAGCGGCTGACCGCAACGTAATATC 567
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
185 InValLeuGlyGlyIleAspIleLeuValGlnIle..... 196
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
568 CATGTGTGTAAGACGACGCGCGACGCGTGGCTGTGAAATGTCGCCAA 617
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
197 ...LeuCys.....ProGluGluValLeuValG 205
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
618 TATCGAAACACATGAA.....TTTGGCGGCGCG..... 645
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
205 yIleGluAspLysProGluAlaIleAlaLeuGlyAlaLeuG 222
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
646 .....CATCTGCGCGCG 657
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
222 lylGluArgProTyrArgIleValAlaLeuProThrArgTyrProSerGly 238
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
658 .....TTGAGTGGCGACGACATTCATTTCAT 683
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
239 GlyGluArgGlnLeuIleGlnLeuLeuThrGlyArgGluValPro...Al 254
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
684 CGAGCCAGTCGGCGCGCAATAAACCCTGTGGACCATCAATATCAAGACG 733
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
254 aAspGlyLeuProAlaAspIleGlyIleLeuCysGlnAsnValGlyThrL 271
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
734 TGATTGCTATCGGACGTTTCTGTAACAGCGCGCTGTAATACCGACGCG 783
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
271 euAlaAlaValHisAspAlaValValLeuGlyArgProLeuIleSerArg 287
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
784 GTGGTCCCTGGCGCGCTGCAAGTCAACAACCGCGCTCTTGGGTAC 833
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
288 IleThrThrLeuAlaGlyIleAlaLeuGluArgProMetAsnValGluAl 304
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
834 CGTTTGGTGGCGAAGTGTCTCACTTACCGCG..... 867
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
304 aLeuIleGlyThrProValHisGluLeuLeuAlaPheAlaGlyLeuAlaG 321
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
868 ..GGCAATGGTGTAGCGCGGACAACCGGTGATTTCCGGTTCGGTATG 915
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
321 luGlyArgLeu.....GluArgValLeuMetGlyGly..... 331
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
916 AACGGTCGATTGCACAAGCGCGCATGATTATTGGGACGCTACCCACA 965
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
331 ..... 331
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
966 TCAGATTTCGGTTATCGAAGAAGCGCGCAAGAGCTGTTCGGTGGG 1015
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
332 .....ProMetMetGlyPheA 337
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1016 TTGCGCGCGACGCGGCAAACTACTCCATCAGCGCACCATCTCGGCCAT 1065
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
337 laLeuProAsp..... 340
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1066 TTCCTAAAAACAACCTCTTCAAGTTCCAGCAGCGCGTCAACGCGCGCA 1115
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
341 ...LeuSerValProLeuIleLysThrCysAsnCysLeuLeuAlaGlyAs 356
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1116 CCGCGCCCATGGTACCG.....ATCG 1135
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
356 paLaThrGluLeuProGluProValProAlaMetProCysIleArgCysG 373
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1136 GCATTATGACGCGGTAATGCGGTGGACATCTCGCTGCTTTTG 1185
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
```

```
373 lyAspCysAlaGlnValCysProValSerLeuLeuProGln...GlnLeu 388
1186 CGCATTATATCGTCGCGATACCGACGCGCGAGCTTTGGTTCCTT 1235
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
389 HisPhePheAlaLeuGlyAsp...GluHisGluGlnLeuLeuAlaHisAs 404
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1236 GGAATTGGACGAAGACCTCTTGTGCAGCTTCGTCTGCGCCGGCA 1285
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
404 nLeuPheAspCysIleGluCysGlyAlaCysAlaTyrValCysProSers 421
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
1286 AATACGAATACGCGCCCTGTTCGCGC 1311
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||
421 erIle.....ProLeuValGln 426
:::||||| ::||| ::||| ::||| ::||| ::||| ::|||

seq_name: SwissProt_40.RNFC_BUCAI

seq_documentation_block:
ID RNFC_BUCAI STANDARD; PRT; 473 AA.
AC P57215;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfC.
GN RNFC OR BUI15.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrtosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN= TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
BU Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
transport (By similarity).
CC -!- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
rnfD, rnfE and rnfG (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
RNFC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AP001118; BAB12833.1; .
InterPro: IPR001450; 4Fe4S ferredoxin.
InterPro: IPR001949; Complex1_51K.
Pfam: PF01512; Complex1_51K; 1.
ProSITE: PS00037; fer4; 2.
Electron transport; Iron-sulfur; 4Fe4S-FERREDOXIN; 2.
FT METAL 337 337 IRON-SULFUR (4FE-4S; Complete proteome.
FT METAL 340 340 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 343 343 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 347 347 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 376 376 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 379 379 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 382 382 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 386 386 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 473 AA; 54563 MW; 04E64102F1315AEE CRC64;
```

alignment\_scores: Quality: 131.00 Length: 440

Ratio: 0.618 Gaps: 20  
Percent Similarity: 48.182 Percent Identity: 20.455  
alignment\_block:  
US-09-303-518D-131 x RNFC\_BUCAI  
Align seg 1/1 to: RNFC\_BUCAI from: 1 to: 473

```

127 ATGAATAAAGAGAGTGAAGCCGTCAAAAAGGCCAA...GTGCGTT 173
      :::::::::: :::::::::: :::::::::: ::::::::::
20  LeuArgValIleAsnGluLysValLeuArgGlyGlnProLeuIlePh 36
      :::::::::: :::::::::: :::::::::: ::::::::::
174 TGAAGACAAAAGATCCGCGGTAGTATTACTGCGCGCGCTTCAGGCA 223*
      :::::::::: :::::::::: :::::::::: ::::::::::
36  eserAspAspPheAsn.....ValProValHisAlaProThrSerGlyL 51
      :::::::::: :::::::::: :::::::::: ::::::::::
224 AAATCGCGCTATTACCGTGGCGAAAGCGGTACTTCAGTACGTGCTG 273
      :::::::::: :::::::::: :::::::::: ::::::::::
51  euIleGluAsnIleCysPheAsnSerAspSerIleLysLysAsnIleLys 67
      :::::::::: :::::::::: :::::::::: ::::::::::
274 ATTGCCGTGTAAGGC.....AACGACGA 296
      :::::::::: :::::::::: :::::::::: ::::::::::
68  IleValIleSerProAspTyrLeuAspGlnTrpIleArgLeuAsnProII 84
      :::::::::: :::::::::: :::::::::: ::::::::::
297 AATCGAGTTCGAACGCTACCTGACCTGAAGCGTGGCAAAATTGACGAGC 346
      :::::::::: :::::::::: :::::::::: ::::::::::
84  eLysAspTyrIleLysTyrAlaProGluLysLeuIleLysIle..... 98
      :::::::::: :::::::::: :::::::::: ::::::::::
347 AAAAAGTGGCGCGCAACTGATTCAATCAGGCTTATGAGTGGCTGCGC 396
      :::::::::: :::::::::: :::::::::: ::::::::::
99  .....IleHisGlnSerGlyVal...ValGlyLeuGly 108
      :::::::::: :::::::::: :::::::::: ::::::::::
397 ACCCGTCGTC.....AGAAATCCCTGCGCTAGATCCGAGGC 437
      :::::::::: :::::::::: :::::::::: ::::::::::
109 GlyGlyGlnPheProSerLysLysIleIlePheSerIleAsnArgAl 125
      :::::::::: :::::::::: :::::::::: ::::::::::
438 GTTCGCCATCTCTCGTACCGATGACACCAATCCGCTGGCTGCGGACC 487
      :::::::::: :::::::::: :::::::::: ::::::::::
125 aHisThrLeuIleValAsnAlaValGluSerGluProTyrIleThrSerA 142
      :::::::::: :::::::::: :::::::::: ::::::::::
488 CTACGGTCATCATCAAGAAGCCGCGAGACTTCAAGCGCGCGCTGTG 537
      :::::::::: :::::::::: :::::::::: ::::::::::
142 sPAsnCysLeuIleTyrAsnHisIleSerGluIleLeuIleGlyCysLys 158
      :::::::::: :::::::::: :::::::::: ::::::::::
538 GTATTGAGCGCGCTGACCGACCTAAATCCATGCTGTAAAGCAGCAGG 587
      :::::::::: :::::::::: :::::::::: ::::::::::
159 IleIleCysTrpIleThrLysIleLysThrValLeuIleAlaIleGlnI 175
      :::::::::: :::::::::: :::::::::: ::::::::::
588 CGCAGAGCTGCCGCTC.....GAAATGCTG 613
      :::::::::: :::::::::: :::::::::: ::::::::::
175 uAspAsnIleGlnSerIleSerLysIleGlnHisLeuIleLysAsnLys 192
      :::::::::: :::::::::: :::::::::: ::::::::::
614 CCAATATCGAACACATGAATTTGGCGCGCCCATCCTCGCGGC..... 657
      :::::::::: :::::::::: :::::::::: ::::::::::
192 erLeuPheLysIleCysIleIleLysLysLysTyrProAlaGlySerSer 208
      :::::::::: :::::::::: :::::::::: ::::::::::
658 .....TTGAGTGCACGCACAT.....CA 677
      :::::::::: :::::::::: :::::::::: ::::::::::
209 LysValLeuValLysSerLeuThrGlyLysGluValProHisGlyLysHi 225
      :::::::::: :::::::::: :::::::::: ::::::::::
678 TTTCATCGACCGAGTGGCGCGGAATAAAACCGTGTGGACCATCATTC 727
      :::::::::: :::::::::: :::::::::: ::::::::::
225 sSerIleAsp.....IleGly.....TyrLeuIlePheAsnValA 237
      :::::::::: :::::::::: :::::::::: ::::::::::
728 AAGACGTGATTCATCGACCGCTTTGTCGTAAACAGCGCGCTCTGAATACC 777
      :::::::::: :::::::::: :::::::::: ::::::::::
237 laThrIlePheSerIleLysArgAlaIleIleAsnGlyLysProLeuThr 253
      :::::::::: :::::::::: :::::::::: ::::::::::
778 GAGCGCGTGGTTCCTTGGCGCGCTGCAAGTCAACAAACCGCGCTCTT 827
      :::::::::: :::::::::: :::::::::: ::::::::::
254 GluArgValValThrLeu.....MetSerAspLysAsnLeuLe 266
      :::::::::: :::::::::: :::::::::: ::::::::::
828 GCGTACCGTTTGGTGGCGAGGTCTCA.....CTTA 862

```

```

266 uSerGlyAsnPheTrpValArgIleGlyThrProIleLysTyrPheLeu 283
      :::::::::: :::::::::: ::::::::::
863 CCGCGCGCGAATTGGTTGACGCGGACAAACCGGTGATTTCGGGTTCGTA 912
      :::::::::: :::::::::: ::::::::::
283 hrSerAsnLysLeuLysGlnSer.....PheIleAlaSerValTyr 296
      :::::::::: :::::::::: ::::::::::
913 TTGAACGGTGGCATTTGCACAAGCGCGCATGATTATTGGGACGCTACCA 962
      :::::::::: :::::::::: ::::::::::
297 LeuGlyGly.....ProPheMetGlyLysLysII 306
      :::::::::: :::::::::: ::::::::::
963 CAATCAGATTTCGGTTATCGAAGAAGCGCGCAGCAAGAGCTGTCGGCT 1012
      :::::::::: :::::::::: ::::::::::
306 eAsnAsnLeuAsn..... 310
      :::::::::: :::::::::: ::::::::::
1013 GGGTTGGCGCGCAGCGCAAAATCTCCATCAGCGCACCACTCTCGGC 1062
      :::::::::: :::::::::: ::::::::::
311 .....HisSerIleLeuLysLysThrAsnSer 319
      :::::::::: :::::::::: ::::::::::
1063 CATTTCTAAAAACAACCTCTTCAAGTTCCAGCACAGCCGCTCAACGCGG 1112
      :::::::::: :::::::::: ::::::::::
320 IlePheIleThrHisLysLysGluLysAsnGluSerIle.....Se 333
      :::::::::: :::::::::: ::::::::::
1113 CGACGCGCGCATGTCACCGATCGCCTTATGACGCGGTAAATGCCGTTGG 1162
      :::::::::: :::::::::: ::::::::::
333 rGluLysThrCysIleArgCysGlyTyrCysSerTyrValCysProValA 350
      :::::::::: :::::::::: ::::::::::
1163 ACATCTCGCTTACCTGCTT.....TTGCGCGATTAAATCGTCGCGC 1203
      :::::::::: :::::::::: ::::::::::
350 snLeuLeuProGlnGlnLeuTyrTriptyrIleLysAsnLysAsnHisVal 366
      :::::::::: :::::::::: ::::::::::
1204 GATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAAATTTGGACGAGAAGA 1253
      :::::::::: :::::::::: ::::::::::
367 GlnThrLysLysHisTyrValLeuAspCysIleGlu..... 378
      :::::::::: :::::::::: ::::::::::
1254 CCTCGCTTTGTGACGCTTCTGTCGCGCGCAATACGAATACGCGCGCGC 1303
      :::::::::: :::::::::: ::::::::::
379 .CysLysAlaCysGluLysValCysPro.....SerTyrIleProL 392
      :::::::::: :::::::::: ::::::::::
1304 TGTTCGCGCAAGTCTGGAA 1323
      :::::::::: ::::::::::
392 euValLysTyrPheIleGln 398
      :::::::::: ::::::::::
seq_name: SwissProt_40:N121_RAT
seq_documentation_block:
ID N121_RAT STANDARD; PRT; 1199 AA.
AC P52591;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
GN NUP121.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RA MEDLINE=93328754; PubMed=8335683;
RX Hallberg E.; Wozniak R.W.; Blobel G.;
RT "An integral membrane protein of the pore membrane domain of the
RL J. Cell Biol. 122:513-521(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE.
CC -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS

```









```

|||||: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
42 ProAlaLeuIleArgLeuProLeu.....GlnGlnHisIleGlyAlaG1 56
120 CCCTCGATGATAAACAAGGAGGTGAAGCGGTCAAAAAGCCCAAGTGC 169
: : : : : |||: : : : : ||| : : : : : ||| : : : : :
56 ualaGluProIleValIysArgAspLeuValLeuLysGlyGlnLeuI 73
170 TGTGTAAGACAAAAGAATCCGGCGGTAGTATTACTGCGCCGGCTTCA 219
: : : : : |||: : : : : ||| : : : : : ||| : : : : :
73 leaAlaLysAlaArgGlyProLeuSerAlaAsnIleHisAlaProThrSer 89
220 GCAAAATCGCGGTATT.....CACCGTGCGGAAAA 251
|||||: : : : : ||| : : : : : ||| : : : : :
90 GlyArgValIleAlaValGlyHisPheValAlaProHisAlaSerGlyLe 106
252 GCGCGTACTTCAGTCAGTCGATTGCGGTGAAGCAACGACCA..... 297
|||||: : : : : ||| : : : : : ||| : : : : :
106 uproVal....ProThrIleThrIleArgProAspGlyGluAspLysTrpG 122
298 .....ATCGAGTTCCGACCGCTACGTACCTGAAGCGCTG 330
122 lyProHisLeuProArgLeuArgProGluAsnAlaAlaProGluGluIle 138
331 GCAAAATTCAGCAGCGAAAAGTCGCGCAACCTGTATTCAATCAGCGTT 380
|||||: : : : : ||| : : : : : ||| : : : : :
139 Ala.....AlaGlnValAlaAlaGlyIleValGlyMetGlyG1 152
381 ATGGACT.....GCGGTTCGCCACCCGTCGCTTCAGCAAAATCCCTG 421
|||||: : : : : ||| : : : : : ||| : : : : :
152 yAlaThrPheProSerAlaValLysLeuAsnLeuArgAlaLysTyAspL 169
422 CCGTAGATCGCGCCGTCGCCATCTTCCTCAATCGGATGGACCAAT 471
|||||: : : : : ||| : : : : : ||| : : : : :
169 euThr.....ThrLeuIleIleAsnGlyAlaGluCysGlu 180
472 CCGGTGCTGCGCACCCCTAGGTCATCATCAAGAAGCCGCGCAAGACTT 521
|||||: : : : : ||| : : : : : ||| : : : : :
181 ProTyrLeuThrCysAspAspArgLeuMetArgGluArgAlaGluGluI 197
522 CAACGCGCGCTGTGTATTGAGCGCGCTGACCGCAACGTAAATCCATG 571
|||||: : : : : ||| : : : : : ||| : : : : :
197 eAlaAspGlyIleGlyIleMetAlaArgAlaLeuGlyValLys...GlnV 213
572 TGTGTAAGCAGCAGCGCGCAGCTGCTGCTGAAATGCTGCC..... 615
|||||: : : : : ||| : : : : : ||| : : : : :
213 alPheValAlaIleGluSerAsnLysProGlnAlaIleGluAlaMetThr 229
616 .....AATATCGAAACACATGAATTTGGCGG 641
230 ArgTyrAsnArgAlaLeuGlyTyrThrPheLysIleHisValValProTh 246
642 CCGCATCTCCGCGCTTGAGTGCACGCACATTCATTTTCAGGCGCAG 691
|||||: : : : : ||| : : : : : ||| : : : : :
246 rGlnTyrPro.....MetGlySerGluLysHisLeuValLysMetI 260
692 TCGCGCGGAATAAACCC.....GTG 711
: : : : : ||| : : : : : ||| : : : : :
260 leThrGlyGlnGluThrProAlaArgAlaLeuThrAlaAspLeuGlyVal 276
712 TGGACCATCAATTCACACAGCTGATTGCTATCGGACGTTGTCGTAA 761
: : : : : ||| : : : : : ||| : : : : :
277 valValHisAsnIleAlaThrAlaHisAlaValHisLeuAlaValArgTy 293
762 AGCCGCTCTGAATACGACGCGGTGCTGCTTGGCGCGCTTCGAAGTCA 811
|||||: : : : : ||| : : : : : ||| : : : : :
293 rGlyGluProLeuIleAlaArgThrValThrValSerGlyHisGlyIleA 310
812 ACAACCGCGCTCTTGCTACCGTTTGGGTGCGAAGTGTCTCAACT 861
: : : : : ||| : : : : : ||| : : : : :
310 rgArgProAlaAsnLeuArgValLeuIleGlyThrProValSerGluIle 326
862 ACCGCC.....GGCAATTGTTGACCGCGGACACCCCGTATTTCGG 905
|||||: : : : : ||| : : : : : ||| : : : : :

```

```

327 IleAlaHisCysGlyGlyPheThrGluProAspArgLeuLeuLeuG1 343
906 TTCGGTATTGAACGGTGGCGATTGCACAAGCGCGCATGATTATTGGGAC 955
: : : : : ||| : : : : : ||| : : : : :
343 yGlyProMetMetGlyMetProIleGln..... 352
956 GCTACCAACAATCAGATTTCGGTTATTCGAAGAAGCCGCGCAAGAGCTG 1005
: : : : : |||: : : : : ||| : : : : :
353 .....AsnProArgValProValValLysGly..... 361
1006 TTCGGCTGGTTGCGCGCGCAGCCGACAAATACTCCATCAGCGCCACC 1055
361 ..... 361
1056 TCTGGGCOATTTCCTAAAAAACACTCTTCAAGTTCCAGCACAGCGCTCA 1105
|||||: : : : : ||| : : : : : ||| : : : : :
362 .....ThrAsnGlyIleLeuAlaLeuThrAlaAlaGluT 373
1106 ACGGCGCGCGCGCGCATG.....GTACCGATCGGCACCTTATGAGCGC 1149
: : : : : ||| : : : : : ||| : : : : :
373 hrProGluAlaLysThrMetProCysIleArgCysGlyArgCysValGln 389
1150 GTAATGCGGTGGACATCCCTACCTGCTGCTTTTTCGCGGATTAACTCT 1199
|||||: : : : : ||| : : : : : ||| : : : : :
390 GlyCysProValGlyLeuThrProPheGluLeuAsnAlaArgIleHisAl 406
1200 CCGGATACCCAGCAGCGCGCGCGCTTTGGGT.....TGCTTGGCAAT 1240
: : : : : ||| : : : : : ||| : : : : :
406 aGlyAspLeuGluGlyAlaAlaLysValGlyLeuMetAspCysLeuAla. 422
1241 TGGACGAAGAAGACCTCGCTTTTGTGAGCTTCGCTGCGCGGCAATATAC 1290
: : : : : |||: : : : : ||| : : : : :
423 .....CysGlyCysSerTyrAsnCysProAlaAsnLeu 434
1291 GAATACGCGCGCGCTGTGCGC 1311
|||||: : : : : ||| : : : : : ||| : : : : :
435 .....ProLeuValGln 438
seq_name: SwissProt_40:RNFC_EC057
seq_documentation_block:
ID RNFC_EC057 STANDARD; PRT; 740 AA.
AC P58324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR Z4636 OR ECS2338.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

```







```

278 rLeuArgAspMetProAsnIleAsnValArgGlyAspPheAlaAsnL 295
923 CGATT..... 927
295 ysileGlnPheMethHisAsnIleAsnIleGlyIleAlaValAspThrPro 311
928 .....GCACAGCGCGCATGATTATTGGGACG 956
312 AsnGlyLeuMetValProValIleIleGlyAla...AspHisLeuSerVa 327
957 CTACACAAATCAGATTCCGTTATCGAAGAAGCGCGCAGCAAAAGAGTGT 1006
327 lPheGluIleAlaIleIleSerGlu..... 336
1007 TCGGCTGGTTCGCGCGCAGCGGACAAATACTCATCAGCGCAC... 1053
337 .....LeuAlaAsnIleAlaIleGlyLeuThrArgAlaGlu 350
1054 .....ACTCTCGGCATTTCCTAAACAAACAACTCTT 1085
351 MetThrGluAlaThrPheThrValSerAsnPheGlySerValGlyLeuAs 367
1086 CAAGTTACGACACCGCTACAGCGGCGCGCGCATGTCACGATCG 1135
367 pTyrAlaThrProIleIleAsnSerProGluSerAlaIleLeuGlyValG 384
1136 GCACCTATGAGCGC..... 1149
384 lYThrMetSerGlnThrProLeuTyrIleAsnGlyLeuGlnLysArg 400
1150 ...GTAATGCGTGGACATC 1167
401 PheIleMetProLeuSerMet 407
seq_name: SwissProt_40: FASL_SCHPO

```

```

seq_documentation_block:
ID FASL_SCHPO STANDARD; PRT: 2073 AA.
AC Q9UUG0; P78799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid synthase subunit beta (EC 2.3.1.86) [Includes: 3-
DE hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);
DE Enoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);
DE carrier-protein] acetyltransferase (EC 2.3.1.38); [Acyl-
DE protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase
DE thioesterase (EC 3.1.2.14)].
GN FASL OR SPAC926.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SUBUNIT.
RX MEDLINE=98359895; PubMed=9693066;
RA Niwa H., Katayama E., Yanagida M., Morikawa K.;
RT "Cloning of the fatty acid synthetase beta subunit from fission yeast,
RT coexpression with the alpha subunit, and purification of the intact
RT multifunctional enzyme complex."
RL Protein Expr. Purif. 13:403-413(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1725-2073 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;

```

```

RT Identification of open reading frames in Schizosaccharomyces pombe
RT cDNAs."
RL DNA Res. 4:363-369(1997).
CC -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN]
CC ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID
CC SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND
CC 3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.
CC -!- CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =
CC LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+)
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-
CC 2,3-dehydroacyl-[acyl-carrier protein] + NADH.
CC -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC carrier protein] + oleate.
CC -!- SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF
CC TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AB010274; BAA36384.1; -
DR EMBL; AL110469; CAB54157.1; -
DR EMBL; D89148; BAA13810.1; -
DR InterPro; IPR001227; Acyltransf_domain.
DR InterPro; IPR003965; Fatty_acid_synth.
DR InterPro; IPR002539; Maoc_dehydratas.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF01575; Maoc_dehydratas; 1.
DR PRINTS; PR01483; FASYNTHASE.
DR Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;
DR Transferrase; Hydrolase; Lyase; NAD; NADP.
KW DOMAIN 1 459 ACETYL TRANSFERASE.
KW DOMAIN 470 858 ENOYL REDUCTASE.
KW DOMAIN 1155 1644 DEHYDRATASE.
KW DOMAIN 1645 2073 MALONYL/PALMITOYL TRANSFERASE.
KW ACT_SITE 270 270 ACETYL TRANSFERASE (BY SIMILARITY).
KW ACT_SITE 1361 1361 DEHYDRATASE (POTENTIAL).
KW ACT_SITE 1828 1828 MALONYLTRANSFERASE (BY SIMILARITY).
KW CONFLICT 222 222 P -> R (IN REF. 1).
SQ SEQUENCE 2073 AA; 230559 MW; D928270838E7C7C4 CRC64;

```

```

alignment_scores:
Quality: 114.50 Length: 461
Ratio: 0.556 Gaps: 25
Percent Similarity: 44.685 Percent Identity: 21.475
alignment_block:
US-09-303-518d-131 x FASL_SCHPO ..
Align seg 1/1 to: FASL_SCHPO from: 1 to: 2073
217 TCAGCAAAATCGCGCTATTACCGTGGCGGCGGCGGCGGTACTTTCAGTC 266
:::|||||::: ::::|::| ::::|::|
948 ThrGlyLysLeu.....MetArgIleGluArgPheThrGlnAs 962
267 AGTC.....GTGATTGCGGTTGAAGGC.....AACGACGAAA 298
:|:::|:::|:::|:::|:::|:::|
962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnAspProt 979
299 TCGAGTTCGAACGCTACGTACCTGAAGCGGTGCGCAAAATTGAGCAGCGAA 348

```





DR PROSITE; PS00370; PEP\_ENZYMES\_PHOS\_SITE; 1.  
 DR PROSITE; PS00742; PEP\_ENZYMES\_2; 1.  
 KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.  
 FT MOD\_RES 441 441 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 809 816 POLY-GLU.  
 SQ SEQUENCE 819 AA; 90652 MW; E09486ABB5270CEA CRC64;

## alignment\_scores:

Quality: 113.00 Length: 475  
 Ratio: 0.467 Gaps: 20  
 Percent Similarity: 50.947 Percent Identity: 20.421

## alignment\_block:

US-09-303-518D-131 x PPSA\_PYRAB

Align seg 1/1 to: PPSA\_PYRAB from: 1 to: 819

```

31 CCCATCGGGGCGACACCGGAGCAAGTCATTATGAC..... 66
|||||.....
235 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 251
67 .GGCCCGCCATTACCGCAAGTCGGTTCCTGGCGAAGATATGTCGCA 115
|||||.....
251 uGlyAlaValSerGlyAlaValThrProAspGluTrpIleValG 268
116 TGCCTCCCTCGATGAAATCAAGGAAGGTGAAGCCGTCAAAAGGCGCAA 165
|||||.....
268 LuLysGlyThrTrpLysIleLysGlu...LysValIleAlaLysLysGlu 283
166 GTGCTGTTGAGACAAAGAATCCGGCGGTAGTATTACTGCGCGCGC 215
|||||.....
284 ValMet.....ValIleArgAsnProGlu 291
216 TTCAGGCAAA.....ATCGCGCTATTACCGGCGC..... 246
|||||.....
291 uThrGlyLysGlyThrValThrValLysValAlaGluTrpLeuGlyProG 308
247 .....GAAAGCGGTACTT.....CAGTCAGTCGTGATTGCC 279
|||||.....
308 LuTrpValGluLysGluValLeuThrAspGluGluIleIleGluValAla 324
280 GTTGAAGCAACGACGAAATCGAGTTCGACGCTACGTACCTGAACGCT 329
|||||.....
325 LysMetGlyGlnLysIleGluGluHisTyrGlyTrpProGlnAspIleG 341
330 GGCATAATTGACGAGCAAAAGTGGCGCGCAACCTGATTCAATCAGGCT 379
|||||.....
341 uTrpAlaTyrAspLysAspGlyLysLeuTyrIleValGlnSerArgp 358
380 TATGGACTGCGTTCGACCCGTCGCTTCAGCAAAATCCCTGCGGTAGAT 429
|||||.....
358 roValThrThrLeuLysGluThrThrThrGluGluValGluVal... 373
430 GCGAGCGGTTCGCCATCTTCGTCATGCGATGCGATGACACCAATCCG...CT 476
|||||.....
374 ...GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyI 389
477 GGCTGCGGACCCCTACGGTTCATCATCAAGGAAGCGCGGAGACTTCAAA 526
|||||.....
389 eGlyAlaGlyArgValValValIlePheAspAlaSerGlu..... 402
527 GCGGCTGTGTTGATTCAGCGCGCTGACCGAAGCTAAATCCATGTGTGT 576
|||||.....
403 .....IleAspLysLysGluGlyAspValLeuValThr 414
577 AAAGCAGCAGCGCGACAG...GTGCGGTCT...GAAATGCTGCCAATAT 620
|||||.....
415 ThrMetThrAsnProAspMetValProAlaMetLysArgAlaAlaIle 431
621 CGAAACACATGAATTGGGGCGCGCGCATCTGCGCGCTTGGAGTGGCAGC 670
|||||.....
431 eIleThrAspGluGlyGlyArgThrSerHisAlaAlaIleValSerArgC 448

```

```

671 ACATTTCATTTCATCGAGCCAGTCGGCGCGCAATAAAACCGTGGACCATC 720
|||||.....
448 LuLeuGlyIleProAlaValValGlyThrLysGluAlaThrLysLysLeu 464
|||||.....
721 AATTATCAAGACCGTGGTTCATCGCA..... 747
|||||.....
465 LysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuValTyrLy 481
748 .....CGTTTTCGTACAGCGCGTCTGTAATACCGAGCGCGTGG 787
|||||.....
481 sGlyIleValLysSerLeuValGluLysLysLysGluAlaAlaA 498
788 TTGCTTGGCGCGCTGCAAGTCAACAACCGCGCTCTTGGTACCGT 837
|||||.....
498 IaAlaProGlyAlaAlaValAlaAlaProLeuValThrGlyThrLeu 514
838 TTGGTTCGGAAGGTGTCTCAACTTACCGCGCGCAATTGTTGACGCGGA 887
|||||.....
515 ValLysValAsnValSerMetProGluValAlaGlu..... 526
888 CAACCGCGGTATTTCGGTTCGGTATTGAACGGTGGATTGCACAGGCG 937
|||||.....
527 ...ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuAlaArgAlaG 542
938 CGCATGATTATTGGGACCGCTACCAATCAGATTTCGGTTCATCAAGAA 987
|||||.....
542 LuHisMetIleLeuSerIleGlyGlnHisProValLysPheIleLysGlu 558
988 GCGCGCAGCAAGAGCTGTTCGGTTCGGTTCGGCGCGCAGCGCGCAATA 1037
|||||.....
559 GlyLysGluGluLeuValGluLysLeuAlaGluGlyIleGluLysVa 575
1038 CTCCATCAGCGCGCACCTCTCGGCGCATTT..... 1067
|||||.....
575 IaAlaAla...AlaPheTyrProArgProValTrpTyrArgThrLeuAspAl 591
1068 .CCTAAAAACCAACTCTTCAAGTT.....CACGACAGCGCTCAAC 1107
|||||.....
591 aProThrAsnGluPheArgGluMetProGlyGlyAspGluProGluG 608
1108 GCGCGGACCGCGCATGTCACCGATCGGCTTATGAGCGCGTAAATGCC 1157
|||||.....
608 LuArgAsnProMetLeuGlyTrpArgGly.....IleArgArgGly 621
1158 GTTGACATCTCTGCTACCTTCTTTCGGCGATTATTCGTCGCGGATA 1207
|||||.....
622 LeuAspGlnProGluLeuLeuArg...AlaGluPheLysAlaIleLysLy 637
1208 CCGACAGCGCGCAGGCTTTCGGTTCGTGGAAATTGGACGAAGAGACCTC 1257
|||||.....
637 sValValGluLysGlyTyrAsnAsnIleGly..... 647
1258 GCTTTGTCGAGCTTCGTCTG.....CCGGGCAATACGAATACGG 1298
|||||.....
648 ....ValMetLeuProLeuValSerHisProGluGlnIleArgLysAla 662
1299 CCGCGTGTGGCAAGTGGTGA 1322
663 LysGluIleAlaArgSerValGly 670
seq_name: SwissProt_40:CHLN_SYNP7
seq_documentation_block:
ID CHLN_SYNP7 STANDARD; PRT; 466 AA.
AC PS4208;
DR 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit N
DE (EC 1.18.-.-) (LI-POR subunit N) (DPOR subunit N).
GN CHLN OR FRXC.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).

```

OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1140;

RN [1]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=95388763; PubMed=7659748;

RX Ronen-Tarazi M., Lieman-Hurwitz J., Gabay C., Orus M.I., Kaplan A.;

RT "The genomic region of rbcL in Synechococcus sp. PCC 7942 contains

genes involved in the ability to grow under low CO2 concentration and

in chlorophyll biosynthesis."

RL Plant Physiol. 108:1461-1469(1995).

CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of

prochlorophyllide (Pchl) to form chlorophyllide a (Chlide)

CC (By similarity). This reaction is light-independent.

CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.

CC -!- SUSUNIT: Prochlorophyllide reductase is thought to be composed

of three subunits; chlL, chlN and chlB. Could form a

heterotetramer of two chlB and two chlN subunits (By similarity).

CC -!- SIMILARITY: BELONGS TO THE BCHN / CHLN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X67694; CAA47924.1; -

DR Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase.

KW SEQUENCE 466 AA; 51540 MW; 9B1A51305045574C CRC64;

SQ

# alignment\_scores:

Quality: 112.50 Length: 471  
 Ratio: 0.577 Gaps: 25  
 Percent Similarity: 41.401 Percent Identity: 20.807

## alignment\_block:

US-09-303-518D-131 x CHLN\_SYN7

Align seg 1/1 to: CHLN\_SYN7 from: 1 to: 466

7 AAATCAAAAAGGCTCTAAATCTGCCATCGCGGGCAGACCGGAGCAAGT 56

92 GlnileLysArgAspArgAsn.....ProSerVal11 102

57 CATTTATGAGCGCGCGCCATTACCGAAGTCGGTTCCTTGGCGAGAAAT 106

102 eValTriPileGlyThrCysThrThrGluIleLysMetAspLeuGlu. 118

107 ATGTCGGCATCGCGCCCTCGATGAAATCAAGGAAGTGAAGCGGTCAA 156

119 .....GlyLeuAlaProLysLeuGluAlaGluIleGlyLeuPro..... 131

157 AAAGGCCAAGTGCTGTTTGAAGCAAAAAGAAATCCGGCGCTAGTATTAC 206

132 .....IleValVa 134

207 TGGCGCGGTTTCAGGCAAAATCCCGCTATTCACCGTGGCGAAAGCGCG 256

134 lAlaArgAlaAsnGlyLeuAspTyrAlaPheThrGlnGlyGluAspThrV 151

257 TACTTCAGTCAGTCGTGATTGCCGTTTGAAGCAACGACGAAATCGAGTTC 306

151 alLeuAlaAlaMetAlaAlaArg..... 158

307 GAACGCTACGTACTGAGCGCTGGCAAAATTTGACGACGAAAGTGGCG 356

159 .....CysProGluAlaAlaThrSerGluAlaAspGlnGlnGluAr 172

357 CCGCAACCTGATTCATCAACGAGCTTATGACTGCGCTTCGCACCCGTCCT 406

172 gThrAsnAlaIleGln.....ArgLeuLeuGlnP 182

407 TCAGCAAAATCCCTGCGCTAGATCCCGAG..... 435

182 heGlyLysSerProAlaAlaGlnGlnProAlaSerSerLysHisPro 198

436 CCGTTCGCGCATCTTCGTCATGCGATGACACCAATCCGCTGGCTCGGA 485

199 ProLeuIleLeuPheGlySerValProAsp.....ProValAlaThrGl 213

486 CCCTACGCTCATCATCAAGAAAGCGCGCAAGACTTCAAAACGCGGCTGT 535

213 nLeuThrIle.....GluLeuAlaLysGlnGlyIleT 224

536 TGGTATTGAGCGGCTGACCGAAGCACTAAATPCCATGTGTGTAAAGACGA 585

224 hrValSerGlyTriLeuProAlaLysArgTyr..... 234

586 GCGCGAGACGTGGCG.....TCTGAAATGCTGCCAATATCGAAACACA 629

235 ...ThrGluLeuProValIleAlaGluGlySerTyrAlaIleGlyLeuAs 250

630 TGAATTT..... 636

250 nProPheLeuSerArgThrAlaThrThrLeuMetArgArgLysCysL 267

637 .....GGCGCGCGCATCTGCGCGCTTGTAGTGGCAGCAC..... 672

267 ysValIleGlyAlaProPheProIleGlyProAspGlySerArgAlaTrp 283

673 .....ATTCAITTCATCGACGACGCTCGC..... 696

284 lIleGluLysIleCysSerValLeuGluIleGluProGlnGlyLeuAlaGl 300

697 .GCGAATAAAACCGGTGGACCATC.....AATTATCAAGACGTGATTG 739

300 uArgGluAlaGlnValTrpAspSerIleGluAspTyrArgGlnLeuValG 317

740 CTATCGGACGTTTGTTCGTAAACAGGC.....CGTCTGAATACC 777

317 luGlyLysGlnValPheMetGlyAspAsnLeuTrpGluIleSerLeu 333

778 GAGCGCGTGTTCCTTGGCGGCGCTG.....CAAGTCAACAAACC 818

334 AlaArgPheLeuValArgCysGlyMetArgCysProGluIleGlyIlePr 350

819 GCGCTCTTTCGTACCGTTCGCGGCGAAGGTCTCAACTTACCGCC. 867

350 oTyrLeuAspArgArgTyrLeuGlyAlaGluLeuAlaMetLeuGluAlaT 367

868 .....GGCGAATGTTGTTGAC....CGGAC 888

367 hrCysGlnSerMetGlyValProLeuProArgLeuValGluLysProAsp 383

889 AACCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 938

384 Asn..... 384

939 GCATGATTATTTGGGACGCTACCAATCAGATTTCCTTATCGAAGAAG 988

385 .....TyrAsnGlnLeuGlnArgIleGlu..... 392

989 GCGGACGAAAGAGTGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 1038

393 .....AlaLeuGlnProAsp..... 397

1039 TCCATCAGCGCACCTCTCGGCCATTTCCTTAAATAAAACAAATC..... 1083

398 ...LeuValIleThrGlyMetAlaHis.....AlaAsnProLeuGluAl 411

1084 .....TTCAAGTTCACGACGCGTCAACG 1108

411 aArgGlyIleSerThrLysTrpSerValGluPheThrPheAlaGlnIleH 428



237 AATAGCGCGGATTTTGGCTGAAGCCGCGCAGTAATATCTACGCCGGAT 188  
 552 .....LeuProGluLysGlnAspGlyAlaThrSerProGlyA 564  
 187 TCTTTTGTCTCAACAGCAGCTGGGCTTTTGGCTTTCAGGCTTCACCTTCC 138  
 564 laLeuLeuProAlaSerSerThrThrSerPhePheTyrSerAsnProHis 580  
 137 TTGATTTTCATCGAGGGCGCATGCCACATATCTTCGCGCAACACGC 88  
 581 AspSerLeuValMetAsnThrLeuThrSerPheSerProLeuAsnG 597  
 87 GACTTCGGTAATGCGCGCGCGTCATAAATGACTTGTCTCGGHTGCGCC 39  
 597 nGluThrAlaValGluAlaProSerArqArqThrAspAspGluLeuPro 613

```

seq_name: SwissProt_40:PSPA_PYRFU

seq_documentation_block:
ID   PSPA_PYRFU          PRT;      817 AA.
AC   P42850; Q59672;
DT   01-NOV-1995 (Rel. 32, Created)
DD   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water
DE   dikinase) (PEP synthase).
DE   PSPA OR Pf0043.
GN   GN
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus
NCBI_TaxID=2261;
LN   [1]
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vcl / DMS 3638;
RX   MEDLINE=95129854; PubMed=7828869;
RT   Robinson K.A., Schreier H.J.;
RT   "Isolation, sequence and characterization of the maltose-regulated
RT   mlrA gene from the hyperthermophilic archaeum Pyrococcus furiosus."
RT   Gene 151:173-176(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vcl / DMS 3638;
RX   MEDLINE=95354939; PubMed=7628701;
RA   Jones C.E., Fleming T.M., Piper P.W., Littlechild J.A., Cowan D.A.;
RT   "Cloning and sequencing of a gene from the archaeon Pyrococcus
RT   furiosus with high homology to a gene encoding phosphoenolpyruvate
RT   synthetase from Escherichia coli.";
RT   Gene 160:101-103(1995).
CC   -1- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC   phosphoenolpyruvate + phosphate.
CC   -1- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC   LACTATE ARE USED AS A CARBON SOURCE.
CC   -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collabora-
CC   tion between the Swiss Institute of Bioinformatics and the EMBL outsta-
CC   tion of the European Bioinformatics Institute. There are no restrictions on
CC   use by non-profit institutions as long as its content is in
CC   modified and this statement is not removed. Usage by and for comm-
CC   enties requires a license agreement (see http://www.isb-sib.ch/ann
CC   or send an email to license@isb-sib.ch).
CC
CC   -----
DR   EMBL; U08376; AA81512.1; -
DR   EMBL; X80819; CAA56785.1; -.
DR   HSPB; P22983; IDIK.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers.C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   Prodom; PD000940; PEP_utilizers; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
DR   Transferrase; Kinase; ATP-binding; Phosphorylation.

```

```

FT   MOD_RES      442  442  PHOSPHORYLATION (BY SIMILARITY).
FT   DOMAIN       809  815  POLY-KLU.
FT   CONFLICT     747  747  Q -> K (IN REF. 2).
SQ   SEQUENCE      817 AA;  90485 MW;  5823694CEF13C74BA CRC64;

alignment_scores:
    Quality: 111.00      Length: 489
    Ratio: 0.459        Gaps: 24
    Percent Similarity: 49.489      Percent Identity: 21.268

alignment_block:
US-09-303-518D-131 x PPSA_PYRFU ..
Align seg 1/1 to: PPSA_PYRFU from: 1 to: 817

```

31 CCCATCGCGGCAGACCGGAGCAAGTCATTATTGAC..... 56  
 235 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 251  
 67 GGCCCGGCATTACCAAGTCGCGTGTGCGGAAGAATATGTCGCA 115  
 251 uGlycAlaValValSerGlyAlaValThrProAspGluIryIleValG 268  
 1116 TCGCGCCCTCGATCAAAATCAAGAGGTGAACCGCTCAAAAAGGCCAA 165  
 268 luLysGlyThrTrpLysIleLysGluLysValIleAlaLysLysGluVal 284  
 116 GTGCTGTTGAAGCAAAAAGATCCG.....GGCGTAGTATT 203  
 285 MetValIle.....ArgAsnProGluThrGlyArgGlyThrValMe 298  
 204 TACTGCGCGCGCTTCAGGCAAAATCGCGCTATTATCACCGTCG..... 246  
 298 tVal.....LysValAlaGlu...TyrLeuGlyProGluTr 309  
 247 .....GAAACGCGGTACTT.....CAGTACTGCTGATTCGCCGTGAA 285  
 309 rpValGluLysGluValLeuThrAspGluGlnIleLeuValAlaLys 325  
 286 GGCACGAGCAAAATCGAGTTCGACGCTACGTACCTGAGACCGTCGCA... 333  
 326 MetGlyGlnLysIleGluAspHisTyrGlyTrpProGluAspIleGluTr 342  
 334 .AAATTGACGACGCAAAAAGTGGCGCGCAACCTGTATTCATCAGCGCTAT 382  
 342 pAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgProI 359  
 383 GGACTGCGCTTCGACACCGTCGCTTCAGCAAAATCCTCGCGTAGATGCC 432  
 359 leThrThrLeuLysGluGluAlaThrAlaGlu...GluAlaGluGluVal 374  
 433 GAGCGCTTCGCGCATCTTCGTCAATGCGCATGGACACCAATCCG...CTGGC 479  
 375 GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIleG 391  
 480 TCGCGACCGCTTCGGTCAATCATCAAGAACGCCGCGAA.....GACTTCA 523  
 391 yAlaGlyArgValValValIlePheAspAlaSerGluIleAspLysValL 408  
 524 AACGCGCG...CTGTGTGTATTGAGCGCGCTGACCGCAACGTAATAATCCAT 570  
 408 ysGluGlyAspIleLeuValThrThrMetThrAsnProAspMetValPro 424  
 571 GTGTGTAAGCAGCGCGCAGCGTAGCGTCCCGCTCGAAAATGCT..... 612  
 425 AlaMetLysArgAlaAlaIleValThrAspGluGlyGlyArgThrSe 441  
 613 ....GCCAATATCGNAACACATGAATTTGGCGCGCCGCATCTCGCGCGGT 658  
 441 rHisAlaIleValSerArgGluLeuGlyIlePro..... 453  
 659 TCAGTGGCAGCAGCATTTTCATTTCATCGACGCGAGTTCGCGCGCAATAAAC 708

```

453 ..... 453
709 GTGTGGACCAATCAATATCAAGACGTGATTCCTATCGGAGCTTTGTCGT 758
454 .....CysValValGlyThrLysGluA 461
759 AACAGCGCGTCTGAATACCGACGCGTGTTCCTTGGCGGCGCTGCAA. 807
461 aThrLysLysLeuLysThrGlyMetTyrValThrValAspGlyThrArgG 478
808 .....GTCAACAACCGCGCTCTGCGTACCGTTTGGTGGCGAAGTG 852
478 lyLeuValTyrLysGlyValLysSerLeuValLysLysGluA 494
853 TCTCAACTACCGCGCGGCGAATTCGTCGCGACGACACCGCGTGATTC 902
495 AlalysAlaGluGlyGlyGlnValValAlaGlyAlaProLeuValTh 511
903 CGGTTCG.....GTATTGAACGGTG 922
511 rGlyThrMetValLysValAsnValSerMetProGluValAlaGluArgA 528
923 CGATTGCAACAGCGCG.....CATCAT 945
528 laAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMet 544
946 TATTGGGACGCTACCAACATCATTCGTTATCGAAGAGCGCGCAG 995
545 IleLeuSerIleGlyGlnHisProIleLysPheIleLysGlyLysGI 561
996 CAAGAGCTGTCGCGTGGTGGCGGCGGCGGACGCGCAATATCATCA 1045
561 uGluGluValValGluLysLeuAlaGluGlyIleGluLysValAlaAa. 577
1046 CGGCGACACTCTCGGCGCAATT.....CCTAATA 1074
578 AlaPheTyrProArgProValTyrArgThrLeuAspAlaProThra 594
1075 ACAAACTCTCAAGTT.....CAGCACGCGCTCAACGCGCGCA 1115
594 snGluPheArgGluMetProGlyGlyGluAspGluProGluGluArgAsn 610
1116 CCGCGCGCTGTCACCGTACCGCTATGACGCGTAATGCGGTTGGACA 1165
611 ProMetLeuGlyTrpArgGly.....IleArgArgGlyLeuAspGI 624
1166 TCCTGCTACCTTGTCTTTGCGCGATTATTCGTCGCGGATACCGACACG 1215
624 nProGluLeuLeuArg...AlaGluPheLysAlaIleLysLysValValG 640
1216 GCGCAGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1265
640 lLysGlyTyrAsnAsnIleGly.....Val 648
1266 CAGCTTCGTCGTG.....CCGGGCAATACGAAATACGCGCGCTGT 1306
649 MetLeuProLeuValSerHisProGluGlnIleArgGluAlaLysArgII 665
1307 TCGCAAGAGTGTGGA 1322
665 eAlaArgGluValGly 670

```

seq\_name: SwissProt\_40:HFCL1\_MESAU

seq\_documentation\_block:

ID HFCL1\_MESAU STANDARD; PRT; 2090 AA.  
AC P51611;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Host cell factor c1 (HCF) (VP16 accessory protein) (HFCL1) (VCAF)  
(DE (CPF)).

```

GN HFCL1
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabepu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC SUBCELLULAR LOCATION: Nuclear.
CC DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PCPE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D45419; BAA08258.1; -.
CC HSP; P02751; IFNA.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR001798; Kelch.
CC Pfam; PF00041; fn3; 2.
CC SMART; SM00060; FN3; 2.
CC SMART; SM00060; FN3; 2.
CC Nuclear protein; Repeat.
CC REPEAT 44 89
CC REPEAT 93 140
CC REPEAT 148 194
CC REPEAT 217 265
CC REPEAT 266 313
CC REPEAT 313 313
CC REPEAT 1010 1448
CC REPEAT 1010 1035
CC REPEAT 1072 1097
CC REPEAT 1101 1126
CC REPEAT 1157 1182
CC REPEAT 1295 1320
CC REPEAT 1323 1348
CC REPEAT 1358 1383
CC REPEAT 1423 1448
CC REPEAT 2090 AA; 214942 MW; E495EBBIF2385E17 CRC64;
SQ SEQUENCE

```

alignment\_scores:

Quality: 109.50 Length: 478  
Ratio: 0.534 Gaps: 22  
Percent Similarity: 42.887 Percent Identity: 21.339

alignment\_block:

US-09-303-518D-131/rev x HFCL1\_MESAU ..

Align seg 1/1 to: HFCL1\_MESAU from: 1 to: 2090

1118 CGGTGCGCGCGTGTGACGGCTGTC.....GTGAACCTGAAGAG 1081

363 LysProProProAlaAlaArgValGlnLeuValArgAlaAsnThrAsnSe 379  
1080 TTTGTTTTTTAGGAAATAGCGAGAGTGGTGGCGGTGATGGAGTATTTG. 1032  
379 rleuGluVal...SerTpglyAlaValAlaThraAlaAspSerTyrLeuL 395  
1031 .....TCCGGCTGCGGGCGCAACCCAG 1011  
395 euGlnLeuGlnLysTyrAspIleProAlaThraAlaAlaThraSer 411  
1010 CCGAACAGCTCTTTGCTGGCGCTCTTCGATAACGGAAATCTGATTGTG 961  
412 ProThr.....ProAsnProValProSerValProAla... 422  
960 GTAGGCTCCCAATAATCATCGCGCTGTGCAATCCACCG..... 918  
423 ....AsnProProLysSerProAlaProAlaAlaAlaProAlaValG 438  
917 .....TTCAATACCGAACCGCAATCATCGCGGTG 888  
438 InProLeuThrGlnValGlyIleThrLeuValProGlnAlaAlaAla 454  
887 TCCGGCTCAACCAATTCG.....CGGCGGTAAAGTTGAGACAC 850  
455 ProProSerThrThrThrIleGlnValLeuProThrValProGlySerSe 471  
849 CTTCCACCCCAAAAGGTACGCAAGAGCGCGGT.....TTGTTGA 809  
471 rleSerValProThrAlaAlaArgAlaGlnGlyValProAlaValLeuL 488  
808 CTTGCAGCGCGCCAGGCAACACCGCGTGC.....GTATTCAGACGG 765  
488 ySvalThrGlyProGlnAlaThrThrGlyThrProLeuValThrMetArg 504  
764 CTTGTT.....ACGAACAAAGCTCGATAGCAATCACGCTTGTATAATT 721  
505 ProAlaGlyGlnAlaGlyLysAlaProValThrValThrSer.....le 519  
720 GATGGTCCACAGGTTTATTCCGCGGACTGCGTCGATGAATGATGT 671  
519 uProAlaSerValArgMetValValProThrGlnSerAlaGlnGlyThrV 536  
670 GCGTGCCTCAACCGCGGAGGATGCGGG..... 642  
536 alileGlySerAsnProGlnMetSerGlyMetAlaAlaAlaAla 552  
641 .....CGCCAAATTCATGTGTTCGATATTCG 613  
553 AlaAlaAlaThrGlnLysIleProProSerSerAlaProThrValLeuSe 569  
612 A.....GCATTTTCAGACGGCA 596  
569 rValProAlaGlyThrThrIleValLysThrValAlaValThrProGlyT 586  
595 CGTCTGCGCTGCTGCTTTACACACATGATTTACGTTTCGTCAGCGGG 546  
586 hrThrThrLeuProAlaThrValLysValAlaSerSerProValMetVal 602  
545 CTCATACCAACAGCGCGCTTGAAGTCTTCGCGGCTCTTTGTATGAT 496  
603 SerAsnProAlaThrArgMetLeuLysThrAlaAlaGlnValGlyTh 619  
495 GACGTAAGGTGCGACGC..... 477  
619 rSerValSerSerAlaAlaAsnThrSerThrArgProIleThrValH 636  
476 ....ACGGATGGTGTCCATCGCA..... 456  
636 isLysSerGlyThrValThrValAlaGlnGlnAlaGlnValValThrThr 652  
456 ..... 456

653 valValGlyGlyValThrLysThrIleThrLeuValLysSerProIleSe 669  
456 ..... 456  
669 rValProGlyGlySerAlaLeuIleSerAsnLeuGlyLysValMetSerV 586  
455 .....TTGACGAAGATGCGCAACGCTCGGCATCT 426  
686 alValGlnThrLysProValGlnThrSerAlaValThrGlyGlnAlaSer 702  
425 ACG.....GCAGG 418  
703 ThrGlyProValThrGlnIleIleGlnThrLysGlyProLeuProAlaG 719  
417 GATTTTCCTCAACGGAGGCGGTGCAAGCGAGTCCATAAGCCCTGATTGAA 368  
719 yThrIleLeuLys...LeuValThrSerAlaAspGlyLysProThr.... 733  
367 TCAGGTTGCGCGCACCTTTTCGCTGCTCAATTTTCCAGCGCTTCAGGT 318  
734 .....ThrIleIleThrThrThrGlnAlaSerGly 743  
317 ACGTAGGTTTCGAACCTGATTCGTCGTCCTTCAACGCAATCACGAC 268  
744 AlaGlySerLysProThrIleLeuGlyIleSerSerValSerProSerTh 760  
267 T.....GACTGAAGTACGCGCTTTTCGCCACGCGTGAATAGCGCGCA 227  
760 rThrLysProGlyThrThrThrIleIleLysThrIleProMetSerAlaI 777  
226 TTTTGCCTGAAGCGCGCA.....GTAATACTAGCCCGCGATTC... 186  
777 leIleThrGlnAlaGlyAlaThrGlyValThrSerThrProGlyIleLys 793  
185 .....TTTGTGCTTCAACACGACGACTTG 163  
794 SerProIleThrIleIleThrThrLysValMetThrSerGlyThrGlyAl 810  
162 GCCT.....TTTGTGACGCTTCACCTTCCTTG 135  
810 aProAlaLysIleIleThrAlaValProLysIle 821  
seq\_name: SwissProt\_40.N121\_HUMAN  
seq\_documentation\_block:  
ID N121\_HUMAN STANDARD; PRT; 1229 AA.  
AC Q9Y2N3; O75115; Q9Y4S7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein  
of 121 kDa) (P145).  
GN NUP121 OR KIAA0618.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cordes M., Bauer C., Holmes A.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 243-1229 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=98403880; PubMed=9734811;  
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
RA Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. X.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro."  
RL DNA Res. 5:169-176(1998).  
RN [3]  
RP SEQUENCE OF 1130-1229 FROM N.A.







```

182 LeuThrSerAspThrProLeuIleSerThrMetSerAlaAlaAs 198
1082 .AGTTGTTGTTTGTAGAAATGCCGAGAGTGGTGGCGCTG
      ::::::::::: ||| ::::
198 pAsnValPheSerSerAlaAsnProIleSerAlaSerLeuThrThrA 215
1043 .....ATGAGTATTTGTCGCGCTGGCGGCGCAACCCAGCCG 1008
215 spSerSerGluSerPheAspGlnThrSerThrAlaGlyAlaIleProVal 231
1007 AACAGCTCTTGGCTCGCGGCTTCTTCGATAACGGAAATCTGATTGGTA 958
      :::::::::::
232 GlnSerSer..... 234
957 GCGTCCCAATAATCATCGCGCCTTGTGCAATGCAACGCGTTCATACCG 908
235 .....AlaAspPheSerSers 240
907 AACCGGAATC...ACGCGGTGTCGCGTCAACCAATTCGCGGCGGTA 861
      :: ||||| ::::: ||||| ::::: ||||| :::::
240 erSerGluIleLeuValGlnSerSerAlaAspPheSerProSer... 255
860 AGTTGAGACACCTTCGACACCAACAGCTACGCAAGAGCGCGGTTGTT 811
      ::||| ::||| ::||| ::|||
256 .....SerProThrThr.....AspIleSe 263
810 GACTTGCAGCGCGCCCAAGGCAACGCGCTCGTATTTCAGACGCGCTG 761
      ||| ||||| ::|||
263 xLeuSerAlaAlaProLeuGlnThrSerGluSerSerPheThrThrA 280
760 TTACGACAACGTCGATGCAATACGCTTGTGATAATTGATGCTCCAC 711
      ::||| ::|||
280 laSerAlaAlaLeuProValSerThrAsp..... 290
710 ACGGTTTTATTTCGCGCGGCTGCGTCAATGATGATGTCGCTGCCACT 661
      ::||| ::|||
291 .....ValAspGlySerSerAlaSerProValValSerMe 302
660 CAACCGCGCAGGATGCGCGCGCCCAATTCATGTTTCGATATTGGCAG 611
      ::|||
302 tSerAlaAlaGly.....GlnIleA 309
610 CATTTTCAGACGCGCTCTGCGCTCTGCTTTACACATGATGTTTA 561
      ::||| ::||| ::|||
309 laSerSerSerThrAspAsnProThrMetSerGluThrPheSerLeu 325
560 CGTTCGCTCAGG.....CGGCTCAA 541
      ||| ::|||
326 ThrSerThrGluValAspGlySerAspValSerSerThrValSerAlaLe 342
540 TACCAACAGCGCGGTTTGAAGTCTTCGCGGCTTCTTGTATGATGACCG 491
      ::||| ::||| ::||| ::|||
342 uLeuSerAlaProPheLeuGlnThrSerThrSerAsnSerPheSerIleV 359
490 TAGGGTCGCGCAGCAGCGGATGCTGCCATCGCATGACGAGATGGCG 441
      ::||| ::||| ::||| ::|||
359 alSerProSerValSerPheValProSerGlnSerSerSerAspValAla 375
440 AACGGTCTGCGCTACTACGCGAGGATTTG..... 411
      ||| ::|||
376 .....SerSerThrThrAlaAsnValValSerSerPheSerAspIrl 390
410 .....CTGACGCGGCTGCGAAGCGGATCCATAAGC 377
      ::|||
390 eProProGlnThrSerThrSerglySerVal..... 400
376 CTGATTGAATCAGTTTCGCGGCGACATTTTTCGCTGCTCAATTTTGCACG 327
      ::||| ::||| ::|||
401 .....ValSerValAlaGlnSerAlaSerAlaLeuAlaPheGlnSer 414
326 GCTTCA.....GGTACGTAGCGTTCTGAACTCGATTCGCTTGC 286
      ::|||

```

```

415 SerThrGluValTyrGlyAlaSerAlaSerSerThrMetSerSerLeuLe 431
285 TTCAACG.....GCAATCAGGACTGACTGAAGTACGCGCTTTT 248
      ||||| ::||| |||
431 uSerThrThrSerLeuGlnSerThrThrLeuAspSerSerSerLeuAlaS 448
247 CGCCACCGTGAATAGCGCGATTTTTCCTGAGCGCGCGAGTAAATACT 198
      || ::||| ::|||
448 erSerSerAlaSerSerAspLeuThrAspTyrGlyValSerSerThr 464
197 ACGCCCGGATTTCTTTTGCT.....TCAACACAGCACTTGGCC 160
      ::||| ::|||
465 AlaSerIleProLeuLeuSerAlaSerGluGlnAlaSerThrSerSe 481
159 TTTTTCAGCGCTTCACTTCTTCTTTCATCAGGAGGCGCG..... 117
      ||| ::||| ::|||
481 rPheSerValValSerProSerValSerPheValProSerGlnSerSerS 498
116 .....ATGCCGACATATTCTTCCCAAGC.....AAC 90
498 erAspValAlaSerThrSerAlaProSerValValSerSerPheSer 514
89 GCGACTTCGCTAATGCGCGCGCTCATATAATGACT 54
      ||||| ::||| |||
515 TyrThrSerLeuGlnAlaGlyGlySerSerMetThr 526
seq_name: SwissProt_40:AMVH_YEAST
seq_documentation_block:
ID AMVH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
glucosidase) (1,4-alpha-D-glucan glucosylase).
GN STA1 OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
STAL.",
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
from Saccharomyces cerevisiae.",
RL FEBS Lett. 239:179-184(1988).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
glucose residues successively from non-reducing ends of the chains
with release of beta-D-glucose.
CC -|- SIMILARITY: TO S.POMBE SPBC215.13.
CC -|- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----

DR EMBL; Z38061; CAA86176.1; -;  
DR EMBL; M16164; AAA35014.1; -;  
DR EMBL; M16165; AAA35015.1; -;  
DR EMBL; X13857; CAA32069.1; -;  
DR PIR; B26877; B26877;  
DR PIR; A26877; A26877;  
DR PIR; S48478; S48478;  
DR SGD; S0001458; MUC1.  
DR KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
KW Signal; Multigene family.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.  
FT DOMAIN 210 1367 SER/THR-RICH.  
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1367 AA; 136110 MW; 91C0E2DBD61AA9D CRC64;

alignment\_scores:  
Quality: 107.00 Length: 546  
Ratio: 0.473 Gaps: 19  
Percent Similarity: 41.392 Percent Identity: 19.963

alignment\_block:

US-09-303-518D-131/rev x AMVH\_YEAST ..

Align seg 1/1 to: AMVH\_YEAST from: 1 to: 1367

1337 TCCTCTCAATGTTCCAGCATTTCGCAACAGCGCGCGTATTCGTA 1288  
||||||| :  
891 SerPheSerThrGlyThrValThrProSerSerSerLysTyr..... 905  
1287 TTTGCCCGGCGACGAGCTGCACAAAGGAGGCTTCTTCGTCCAAT 1238  
||||||| :  
906 ....ProGlySerGlnThrGluThrSerValSerSerThrThrGluThrT 921  
1237 CCAAGCAACCAAGCTCGCGCTCGGTA.....TCGCGG 1200  
:  
921 hrleValProThrLysThrThrThrSerValThrThrProSerThrThr 937  
1199 ACGATTAAATCGCGCAAAAGAGTAGGAGGATGTCACACGGCATTC 1150  
||||||| :  
938 ThrileThrThrValCysSerThrGlyThrAsnSerAlaGlyGluTh 954  
1149 GCGCTATAAGTCGGATCGGTACCATGGCGGCTCGCGCGCTTGACGG 1100  
:  
954 rThrSerGlyCysSerProLysThrValThrThrValProThrThr 971  
1099 CTGTCGTGAAGTGAAGAGTTGTTTATAGAAATGGCGAGAGTGGT 1050  
:  
971 hrThrThrSerValThrThr.....SerSerThrThrThrile 983  
1049 CGCGTGATGAGATTTGTCGGCTGGCGGCAACCCAGCGACACCTC 1000  
:  
984 ThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrSe 1000  
999 TTTCTCGGCGCTTCTCGATGAGAACTGTATTGTGTAGCTGCCA 950  
| :  
1000 rGlySerProLysThrileThr..... 1008  
949 AATAATCATGCGCCCTGTGCAATCCACCGCTCAATACCAACCGGAA 900  
:  
1009 .....ThrThrValProCysSerThrSerProSerGluThrAlaserGlu 1023  
900 ..... 900

1024 SerThrThrThrSerProThrThrProValThrThrValValSerThrTh 1040  
899 .....ATCACCGGTGTCGCGTCAACCAATTTCGCGGGGTAAGTT 857  
||||||| :  
1040 rValValThrThrGluThrSerThrThrThrThrProGlyGlyGluThr 1057  
856 GAGACACCTTCGACCCCAACAGGTACGCAAGAGCGCGGTTTGTGACT 807  
||||||| :  
1057 hrThrThrPheValThrLysAsnile...ProThrThrThrLeuThrThr 1072  
806 TGAGGCGG...CCCAAGGCAACACGCGCTCGGTATTTCAGAGCGCTGT 760  
||||||| :  
1073 IleAlaProThrProSerValThrThr.....Va 1082  
759 TACGAACAAACGTCGATAGCAATACGCTTGTGATAATTGATGTCACCA 710  
||||||| :  
1082 lThrAsnPheThrProThrThrThrThr.....T 1093  
709 CGGTTTATTCGCGCGGCTGCTGATGAATGAATGTCGTCGCCACTC 660  
||||||| :  
1093 hrValCysSerThrGlyThrAsnSer.....AlaGlyGlu 1104  
659 AAGCGCGCAGATCGCGCGG..... 639  
1105 ThrThrSerGlyCysSerProLysThrValThrThrValProCysSe 1121  
639 ..... 639  
1121 rThrGlyThrGlyGluThrThrGluAlaThrLeuValThrThrA 1138  
639 ..... 639  
1138 laValThrThrValValThrThrGluSerSerThrGlyThrAsnSer 1154  
638 .....CCAAATTCATG 628  
1155 AlaGlyLysThrThrThrGlyThrThrLysSerValProThrThrTh 1171  
627 TGTTCGATATTCGACGATTTTCAGACGCGACGCTGCGCGCTGCTGCTT 578  
||||||| :  
1171 rValThrThrLeuAlaProSerAlaProValThrProAlaThrAsnAla 1188  
577 TACACACATGATTTACGTCGTCGCGGCTCAATACCAACAGCGCG 528  
:  
1188 alProThrThrThrThrThrGluCysSerAlaAlaThrAsnAlaAla 1204  
527 CGTTTGAAGTCTTCGCGGCTTCTTGTGATGATGACCGTAGGTCGCGCAG 478  
:  
1205 GlyGluThrThrSerValCysSerAlaLysThrThrThrThrThrAla 1221  
477 CAGCGGATGTTGTCCTCCATCGCATTCGCAAGATGCGGACGCGCTCGCAT 428  
:  
1221 rAlaGly.....GluAsnThrAlaP 1228  
427 CTACGCGGAGGATTTGCTGAACGAGCGGTCGAAAGCGCAGTCCATAAG 378  
:  
1228 roSerAlaThrThrProValThrThrAlaThrThrThrVal..... 1242  
377 CCTGATGAATCAGGTTCGCGGCGACTTTTTCGTCGCTCAATTTTGCAG 328  
||||||| :  
1243 .....IleThrThrGluSerSerValGlyThrAsnSe 1253  
327 CGCTTCAGGTACGTACGTCGCACTGATTCGTCGTCGTCGTCGTCGTCG 278  
||||||| :  
1253 rAlaGlyGluThrThrThrGlyThrThrLysSerThrThrThrThrThr 1270  
277 CAATCAGCATGCTACTGAAGTACGCGCTTTTCGCCACGCGTGAATAGCGCG 228  
||||||| :  
1270 yrIleThrThr..... 1273  
227 ATTTTGCCT.....GAAGCGCGCGCACTAAA 202  
:  
1274 LeuIleProGlySerAsnGlyAlaLysAsnThrGluThrValAlaThrAl 1290

```
201 TACTAGCC...GGATCTTTTGT 179
|||||
1290 athrasnProleSerileLysThrThrSerGlnLeuAlaThrAlaAs 1307
178 CTTCAACAGCAGCTGGCTTTTGGAGCGTTCACCTTCCTTGATTTTC 129
|||||
1307 erAlaSerSerValAlaProValValThr...SerProSerLeu..... 1320
128 ATCGAGGGCGCATGCGCATATATCTTCGCAAGCAGCGACTTGGT 79
|||||
1321 .....ThrGlyProLeuGlnSerAlaSerGlySerAlaValAlaLath 1334
78 AATGCCGGCGGCATATAATGACT.....TGCTCCGCTGTCG 41
|||
1334 rTySerValProSerileSerSerThrTyrglnGlyAlaAlaAsnleL 1351
40 CCGGATGCGCAGATTAGACTTTTGTGATTTTAATC 3
|||||
1351 ysValLeuGlyAsnPhMetTrpLeuLeuAlaLeu 1363

seq_name: SwissProt_40:MUC1_MOUSE

seq_documentation_block:
ID MUC1_MOUSE STANDARD; PRT; 630 AA.
AC Q02456;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Episialin).
GN MUC1 OR MUC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=91332029; PubMed=1714452;
RA Spicer A.P., Parry G., Fattori S., Gendler S.J.;
RT "Molecular cloning and analysis of the mouse homologue of the tumor-
RT associated mucin, MUC1, reveals conservation of potential O-
RT glycosylation sites, transmembrane, and cytoplasmic domains and a
RT loss of minisatellite-like polymorphism."
RL J. Biol. Chem. 266:15099-15109(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92068178; PubMed=1958179;
RA Vos H.L., Devries Y., Hilkens J.;
RT "The mouse episialin (Muc1) gene and its promoter: rapid evolution of
RT the repetitive domain in the protein."
RL Biochem. Biophys. Res. Commun. 181:121-130(1991).
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED
CC IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY
CC POLARIZED EPITHELIAL CELLS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL
CC TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M64683; AAA397556.1; -
DR EMBL; U16175; AAA98538.1; -
DR EMBL; M65132; AAA39755.1; -
DR EMBL; M64928; AAA39755.1; JOINED.
```

```
DR EMBL; M77226; AAA39754.1; -.
DR PIR; A39344; A39344.
DR MGD; MGI:97231; Muc1.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW Repeat.
FT SIGNAL 1 20
FT CHAIN 21 630
FT DOMAIN 21 535
FT TRANSMEM 536 556
FT DOMAIN 557 630
FT DOMAIN 42 366
FT REPEAT 42 61
FT REPEAT 62 81
FT REPEAT 82 101
FT REPEAT 102 122
FT REPEAT 123 143
FT REPEAT 144 164
FT REPEAT 165 184
FT REPEAT 185 204
FT REPEAT 205 225
FT REPEAT 226 246
FT REPEAT 247 256
FT REPEAT 257 286
FT REPEAT 287 306
FT REPEAT 307 326
FT REPEAT 327 346
FT REPEAT 347 366
FT DOMAIN 411 526
FT CARBOHYD 125 125
FT CARBOHYD 275 275
FT CARBOHYD 302 302
FT CARBOHYD 335 335
FT CARBOHYD 355 355
FT CARBOHYD 366 366
FT CARBOHYD 408 408
FT CARBOHYD 432 432
FT CARBOHYD 449 449
FT CARBOHYD 508 508
FT CONFLICT 120 120
FT CONFLICT 121 121
FT CONFLICT 138 139
FT CONFLICT 140 140
FT CONFLICT 423 423
FT CONFLICT 506 506
FT CONFLICT 602 602
SQ SEQUENCE 630 AA; 64622 MW; FF57C1B31137C83B CRC64;

alignment_scores:
Quality: 106.50 Length: 456
Ratio: 0.520 Gaps: 21
Percent Similarity: 44.956 Percent Identity: 23.246

alignment_block:
US-09-303-518d-131/rev x MUC1_MOUSE ..
Align seg 1/1 to: MUC1_MOUSE from: 1 to: 630

1337 TCCTTCTCAATGGTTTCAGCAGCTTTGGCAACAGCGCGCGTATTCGTA 1288
|||||
38 SerSerSerLeuAlaSerThrThr..... 45
1287 TTTCGCGCGGAGAGCAGCTTCGCAACAGCGAGGCTTCTTCGTCGAATT 1238
|||||
46 .....ThrProValHisSer.....SerAsnSerAspProA 56
1237 CCAAGCAACCAAGCCTGCGCGTGTGCGTATCGCGCAGCATTAATCG 1188
|||
56 laThrArgProGlyAspSerThrSer....SerProValGlnSerSer 71
```

1187 CCACAAACAGAGGTAGCGAGGATGTCACACGGCAATTACGCGCTCATTAAGT 1130  
 72 ThrSerSerProAlaThrArgAlaProGluAspSerThrSerThrAlaVala 88  
 1137 GCGGATCGGTACCATGGCGCGTGGCGCGCTTACACGGCTGTCGTGAACT 1088  
 88 LleuSerGlyThr.....SerSerProAlaThrThrAlaProValAsnS 103  
 1087 TGAAGAGATTGTTTTTGGAAATGGCCGAGAGTGGTCCGCGTGAATGAG 1038  
 103 eAlaSer.....Ser.....Ser.....Ser 106  
 1037 TATTGTTCGGCTGCGGGCGCAACCCAGCGCAACAGCTCTTTGCTGCGGCC 988  
 107 ProValAlaHisGlyAspThrSerSerProAlaThrThrAlaProValAsnS 123  
 987 TTTCTCGATAACGGAAATCTGATTGTGCTAGCGTCCCAATAAATCATGCG 938  
 123 pSerAsnSerSerProVal.....Ser.....Ser.....Ser 120  
 937 CGCCTTGTGCAATCGCACCGTTCATACCGAACCGGAAATCACGCGGTG 888  
 130 .....ValHisSerGlyThrSerSerAlaAlaThrThrAlaPro 142  
 887 TCGCGCTCAACCAATTCGCGCGGTAGTTGACAGACCTTCGACACCCAA 838  
 143 ValAspSerThrSerSerProValValHisGlyGlyThrSerSerProAl 159  
 837 AACGGTACCAAGAGCGCGGTTTCTGACTTGCAGCGCGCCCAAGGCAA 788  
 159 aThr.....SerProGlyAspSerThrSerSerProAspHisSerS 174  
 787 CACGGCTCGGTATTCAGACGGCTGTTCAGCAACAACGTCCTGATGCA 738  
 174 eThrSerSer.....ProAlaThr...ArgAlaProGluAsp 185  
 737 ATCAGCTCTTGATTAATTGATGTCACACGGTTTTATTTCGCGCGCATGG 688  
 186 SerThrSer.....ThrAlaValLeuSerGlyThrSe 196  
 687 CTCGATGAATGAATGTCGTGCCACTCAAGCGCGCAGGATGCGGCGCG 638  
 196 rSer.....ProAlaThrThrAlaProVal 204  
 637 CAATTCATGCTGTTTCGATATTGGCA..... 612  
 204 aAlaSerThrSerSerProValAlaHisAspAspThrSerSerProAla 220  
 611 ...GCATTTTCAGAC.....GGCAC 595  
 221 ThrSerLeuSerGluAspSerAlaSerSerProValAlaHisGlyGlyTh 237  
 594 GTCTGCGCTGCTGCTTTACACATGGATTTTACGTTTCGTCGTCAGCGCG 545  
 237 rSerSerProAlaThr.....SerProLeuArgA 247  
 544 TCAATACCAACAGCGCGCTTTGAAGTCTTCGCGCGCTCTTTGATGATG 495  
 247 spSerThrSerSerProValHisSerSerAlaSerIleGlnAsnIleuys 263  
 494 ACGGTAGGTCGCCACCCAGC.....GGATTGGTCTCATCGC 457  
 264 ThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValTh 280  
 456 ATTGACGAGATGGCAACGGCTCGGCA..... 429  
 280 rThrThrSerSerAlaLeuGlySerAlaThrSerProAspHisSerGlyT 297  
 428 .....TCTACGGCAGGATTTGCTGAACGACGACGGGTG 396  
 297 hrSerThrThrAsnSerSerGluSerValLeuAlaThrThrProVal 313

[illegible]

CCGCTTGGCAGCACATTT 675

440 alPheThrGlySerProGlyLysTyr 448

seq\_name: SwissProt\_40:ATPB\_TOBAC

seq\_documentation\_block:

ID	ATPB_TOBAC	STANDARD;	PRT;	498 AA.
AC	P00826;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	ATP synthase beta chain (EC 3.6.3.14).			
GN	ATPB			
OS	Nicotiana tabacum (Common tobacco).			
OG	Chloroplast.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.			
OX	NCBI_TaxID=4097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. BRIGHT YELLOW 4;			
RC	Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,			
RA	Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,			
RA	Yamaquichi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M.,			
RA	Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,			
RA	Tohdoh N., Shlmada H., Sugiura M.;			
RT	"The complete nucleotide sequence of the tobacco chloroplast genome:			
RT	its gene organization and expression.";			
RL	EMBO J. 5:2043-2049(1986).			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=84059075; PubMed=6227526;			
RX	Shinozaki K., Deno H., Kato A., Sugiura M.;			
RT	"Overlap and cotranscription of the genes for the beta and epsilon			
RT	subunits of tobacco chloroplast ATPase.";			
RL	Gene 24:147-153(1983).			
RP	[3]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=CV. XANTHI;			
RC	MEDLINE=92390714; PubMed=1387730;			
RA	Avni A., Anderson J.D., Holland N., Rochaix J.-D., Gromet-Elhanan Z.,			
RA	Edelman M.;			
RT	"Fenton's sensitivity of chloroplasts determined by codon 83 of beta			
RT	subunit of proton-ATPase.";			
RL	Science 257:1245-1247(1992).			
CC	-1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON			
CC	GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC			
CC	SUBUNIT.			
CC	-1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +			
CC	H(+)(Out).			
CC	-1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC			
CC	CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE			
CC	SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)			
CC	HAS THREE MAIN SUBUNITS: A, B AND C.			
CC	-1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.			
CC	-1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; Z00044; CAA77360.1; -			
DR	EMBL; K00507; AAA84576.1; -			
DR	EMBL; X61319; CAA43612.1; -			
DR	PIR; A01027; PWN7B.			
DR	HSSP; P00829; LBMF.			
DR	Mendel; 4300; NICTa:atpb1.			
DR	InterPro; IPR003593; AAA.			
DR	InterPro; IPR004100; ATP-synt_ab_N.			



174 .....AlaGlyValGlyLysThrValLeuMetGluLe 185  
726 TCAAGACGTGATGCT.....A 742  
185 uileAsnAsnIleAlaLysAlaHisGlyGlyValSerValPheGlyGlyV 202  
743 TCGGAGCT.....TTGTTCTGTA..... 759  
202 aIGlyGluArgThrArgGluGlyAsnAspLeuMetGluMetLysGlu 218  
760 ACAGGCGCTGTAATACACGAGCGGTG.....GTGCTTGGG 797  
219 SerGlyValIleAsnGluGluAsnIleAlaGluSerLysValAlaLeuVa 235  
798 CGGCTGCAAGTCACAAACCGCCCTCTTCGTACCGTTTGGGTGCGA 847  
235 lTyGlyGlnMetAsnGluProGlyAlaArgMetArgValGly.... 250  
848 AGGTCTCTCAACTTACCGCCGCGAA..... 873  
251 ..LeuThrAlaLeuThrMetAlaGluTyPheArgAspValAsnGluGln 266  
874 .....TTGGTGTGCGGCAAC.....CGCGTCATGTTTCCGTTATC 981  
267 AspValLeuPheIleAspAsnIlePheArgPheValGlnAlaGlySe 283  
909 GGTATTGAACGGTTCGATTCACAA.....GGCGCGCATGATTAT.... 948  
283 rGluValSerAlaLeuLeuGlyArgMetProSerAlaValGlyTyrglnp 300  
949 .....TTGGAGCGCTACCACATCATCATTCGTTATC 981  
300 roThrLeuSerThrGluMetGlySerLeuGlnGluArgIleThrSerThr 316  
982 GAAGAAGCGCGCAGCAAGAGCTTTCGGCTGGGTTCGCGCGAGCGGA 1031  
317 LysGluGlySerIleThrSerIleGlnAlaValTyValProAlaAsp 333  
1032 CAATATCTCC....ATCAGCGGCACCTCTCGGCCATTTCCTAAAAACA 1078  
333 pLeuThrAspProAlaProAlaThrPheAlaHis..... 345  
1079 AACTCTTCAAGTTCACGACGCGTCAACGCGCGCGCGCATCGGTA 1128  
346 .....LeuAspAlaThrThrValLeuSer.....ArgGlyLeuAla 357  
1129 CCGATCGCACTTATGAGCGGTAATGCCGTGGACATCTGCTACCTT 1178  
358 AlaLysGlyIleTyProAlaValAspProLeuAspSerThrSerThr... 373  
1179 GCTTTTCGCGATTATCTCGCGCAT..... 1206  
374 ..MetLeuGlnProArgIleValGlyGluGluHisTyrgluThrAlaGlnA 390  
1207 .....ACGACGCGCGCAG 1221  
390 rgValLysGlnThrLeuGlnArgTyLysGluLeuGlnAspIleIleAla 406  
1222 GCTTTGGTGTGTTGAATGGACGAGAGACCTCGTTTGTGACG... 1269  
407 IleLeuGlyLeuAspGluLeuSerGluGluAspArgLeuValAlaAr 423  
1270 .....TTGCTGTC.... 1278  
423 gAlaArgLysIleGluArgPheLeuSerGlnProPhePheValAlaGluV 440  
1279 .....CCGGCAATAC 1290  
440 alPheThrGlySerProGlyLysTy 448  
seq\_name: SwissProt\_40:PPSA\_PYRHO  
seq\_documentation\_block:

ID PPSA\_PYRHO STANDARD; PRT; 821 AA.  
AC O57830:  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water  
DE dikinase) (PEP synthase).  
GN PPSA OR PH0092.  
OS Pyrococcus horikoshii.  
OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=98344137; PubMed=9679194;  
RX Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +  
CC phosphoenolpyruvate + phosphate  
CC -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND  
CC LACTATE ARE USED AS A CARBON SOURCE.  
CC -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP000001; BAA29161.1;  
DR HSSP; P22983; 1DIK.  
DR InterPro; IPR001121; PEP\_utilizers.  
DR InterPro; IPR002192; PDK\_N\_term.  
DR Pfam; PF00391; PEP\_utilizers; 1.  
DR Pfam; PF02896; PEP\_utilizers; 1.  
DR Pfam; PF01326; PDK\_N\_term; 1.  
DR ProDom; PD000940; PEP\_utilizers; 1.  
DR PROSITE; PS00370; PEP\_ENZYMES\_PHOS\_SITE; 1.  
DR PROSITE; PS00742; PEP\_ENZYMES\_2; 1.  
DR Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.  
KW MOD\_RES 444 444 PHOSPHORYLATION (BY SIMILARITY).  
FT DOMAIN 812 818 POLY-GLU.  
SQ SEQUENCE 821 AA; 90812 MW; A93816D865F8A0BF CRC64;

## alignment\_scores:

Quality: 105.00 Length: 477  
Ratio: 0.438 Gaps: 21  
Percent Similarity: 50.314 Percent Identity: 21.174

## alignment\_block:

US-09-303-518D-131 x PPSA\_PYRHO ..

Align seg 1/1 to: PPSA\_PYRHO from: 1 to: 821

31 CCCATCGCGGCGAGACCGGACGAGCAAGTCATTATGAC..... 66  
|||||.....  
238 ProValThrAsnAsnArgAsnGluIleMetIleAsnAlaSerTrpGlyLe 254  
67 .GGCGCGGCCATTACCGAAGTCGGTTGCGGGAAGAAATATGTCGCA 115  
||| |||.....  
254 uGlyGluAlaValSerGlyAlaValThrProAspGluTyrlleValG 271  
116 TGCGCCCTCGATGAAATCAAGGAGGTGAGCCGTCAGGCAAGGCCAA 165



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; LJ5625; AAB00104.1; -.
DR	InterPro; IPR001374; R3H.
DR	InterPro; IPR000058; ZnF_AN1.
DR	Pfam; PF01424; R3H; 1.
DR	SMART; SM00393; R3H; 1.
DR	SMART; SM00154; ZnF_AN1; 1.
DR	Helicase; ATP-binding; Nuclear protein;
KW	Transcription regulation; Activator.
KW	ATP (POTENTIAL).
FT	NP_BIND 213 220
FT	DOMAIN 250 425
FT	DOMAIN 370 373
FT	POLY-VAL.
FT	SS DNA-BINDING (BY SIMILARITY).
FT	DOMAIN 793 859
FT	GLN/PRO-RICH.
FT	POLY-LYS.
FT	DOMAIN 860 864
FT	DOMAIN 860 864
SQ	SEQUENCE 989 AA; 108439 MW; 9489671E46DA04E CRC64; NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

alignment\_scores:

Quality:	105.00	Length:	473
Ratio:	0.465	Gaps:	27
Percent Similarity:	47.780	Percent Identity:	21.987

alignment block:

US-09-303-518D-131 x SMB2 MESAU

Align seg 1/1 to: SMB2\_MESAU from: 1 to: 989

```

166 GTGCTGTTTGAAGACAAAAG...AATCCGGCGGTACTATTACTGGGCC 212
    ||| ||||| :|||| |::||| |||||::| ::
65 ValThrPheGluProArgLysLeuGlyProValValValLeuProSerAs 81
    .....TCAGCAAAATCGCGCTATTACCGTGGCGAANAAG....252
    ::: ||||| ||||| :|||||::| :|||||::|
81 nSerPheThrSerGlyAspIleValGlyLeuTyrAspAlaAsnGluSer 98
    .....CGCGTACTTCAAG.....CAGTGCTGGT 273
    ..erGlnLeuAlaThrGlyValLeuThrArgIleThrGlnLysSerValThr 114
    274 ATTCCCGGTT...GAAGGCACAGCAC...GAATCGAGTTCGAACGCTACGT 317
    :|||: |||||::|::|::|::|::|::|::|::|
115 ValAlaPheAspGluSerHisAspPheGlnLeuAsnLeuAspArgGluAs 131
    318 ACCTGAAGCGGTGCGAAAAATTGACACACGAAAGAAGTCGCGCGCAACTGA 367
    ||| |||||::|::|::|::|::|::|::|::|
131 nThrTyrArgLeuLeuMetThrLeuLysLysTyrHisSerGlyProAlaSer 163
    412 AAAATC.....CTGCCCGTAGA 428
    |||||::|::|::|::|::|::|::|
164 SerLeuIleAspValLeuLeuGlyGlySerSerProSerProThrThrGl 180
    429 TGCCGAGCCGTGCGCATCTTCGTCAATGCGGATGCACACAAT.....471
    : |||||::|::|::|::|::|::|::|
180 uileProPropheThrPheTyrAsnThrAlaLeuAspProSerGlnLysG 197
    .....471
    197 luAlaValSerPheAlaLeuAlaGlnLysgluValAlaIleHleHisGly 213
    472 ..CCGTGCTGCGCACCTACGGTTCATCATCAAGAAGACCCCGCCAAGA 518
    ||| :|||: |||||::|::|::|::|::|::|
214 ProGlyThrGlyLysThrThrValValGluIleLeuGlnAl 230
    519 CTTCAAACGCGGCTGTTGGTATTGTAGCGCGCTCGACCAAGCTATAAATCC 568

```

230 aValLysGlnGlyLeuLysIleLeu.CysCysAlaProSer.....Asn 244  
 569 ATGTGTGTTAAAGCAGCGCAGCAGGTCCCGCTCTGAAATTCCTGCCCAAT 618  
 245 ValAlaValAspAsnLeuValGlu.....ArgLeuAlaLeu.CysLysL 259  
 619 ATCGAACAACATGAAATTTGGCGGCCGATCCTGCCGCGTTAGTGGCAC 668  
 259 yArgIleLeuArgLeuGlyHisProAlaArgLeuLeuGluSerAlaGln 275  
 669 GCACATTCATTTTCATCGACCGAGTCGGCGCGAATAAAACCGTGGACCA 718  
 276 GlnHisSerLeuAspAlaValLeuAlaArgSerAspAsnAlaGlnIleVa 292  
 719 TCAATTTATCAAGACGTGATGCTATCGACGCTTTGTTGCTAACAGCGCGT 768  
 292 lAlaAspIleArgLysAspIleAspGlnValPheGlyLysAsnLysLysT 309  
 769 CTGAATACCGACGCGGTGGTTCCTTGGCGCGCTGCAAGTCACACAACC 818  
 309 hArgin..AspLysArgLulysSerAsnPheArg.....AsnGluI 322  
 819 GGCCTCTTCGTACCGTTTGT.....GGTCGGAAGGTGT 853  
 322 eLysLeuLeuArgLysGluLeuLysGluArgGluAlaIleValG 339  
 854 CTCACCTTACCGCGCGAATTTGGTTGACGGCGGACACCGGTGATTTC 903  
 339 InSerLeuThAlaAlaAspValValLeuAlaThrAsn..... 351  
 904 GGTTCGGTATTGAACGGTGGCATTCGACAAAGCGCGCATGATTATTTGGG 953  
 352 .....ThrGlyAlaSerSerAspGlyProLeuLysLeuLeuPr 364  
 954 AGCTTACCAACATCAGATTCCTCGTTATCGAAGAA...GGCCGCGACGAAG 1000  
 364 oGluAsnHisPheAspValValValAlaLaspGluCysAlaGlnAlaLeuG 381  
 1001 ACCTGTTTCGGTGGGTT.....GGCCGCGAG..... 1026  
 381 lAlaSerCysTrpIleProLeuLeuLysAlaProLysCysIleLeuAla 397  
 1027 CGGGAACAATACTCCATCACCGCACCACTCTCGGCATTCCTCAAAAA 1076  
 398 GlyAspHisArgGlnLeuProProThrThrIleSerHis..... 410  
 1077 CAAACTCTCAAGTTCACGACAGCGGTACAGGGCGGCGACCGGCCCATGG 1126  
 411 .....LysAlaLeuAlaGlyLeuSerArgSerLeuM 422  
 1127 TACCACCGCACTTATGAGCGCGTAATG..... 1155  
 422 et.....GluArgLeuValGluLysHisGlyAlaGlyAla 433  
 1155 ..... 1155  
 434 ValArgMetLeuThrValGlnTyrArgMetHisGlnAlaIleThrArgTr 450  
 1156 .....CGTTGGACATCCTCGCTACCTGTG. 1179  
 450 pAlaSerGluAlaMetTyrHisGlyGlnLeuThrAlaHisProSerValA 467  
 1180 .....CTTTTGGCGATTAAATC...GTCCGGCATACCGACGCGG 1218  
 467 laGlyHisLeuLysAspLeuProGlyValAlaAspThrGluGluThr 483  
 1219 CAG.....GCTTTGGGTTCGTTGGAAT 1241  
 484 SerValProLeuLeuIleAspThrAlaGlyCysGlyLeuLeuGluLe 500  
 1242 GGACGAAGAC 1254

500 uAspGluGluAsp 504

seq\_name: SwissProt\_40:ILVE\_BACSU

seq\_documentation\_block:

ID ILVE\_BACSU STANDARD; PRT; 362 AA.

AC P39576;

DT 01-FEB-1995 (Rel. 31, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)

DE (BCAT) (Vegetative protein 85) (VEG85).

GN YWAA OR IPA-OR.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RC STRAIN=168;

RA Glaser P., Lubochinsky B., Danchin A.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-57 FROM N.A.

RC STRAIN=168;

RX MEDLINE=95020537; PubMed=7934828;

RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,

RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,

RA Rapoport G., Danchin A.;

RT "Bacillus subtilis genome project: cloning and sequencing of the 97

RT kb region from 325 degrees to 333 degrees."

RL Mol. Microbiol. 10:371-384 (1993).

RN [3]

RP SEQUENCE OF 1-9.

RC STRAIN=IS58;

RX MEDLINE=97443988; PubMed=9298659;

RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,

RA Hecker M.;

RT "First steps from a two-dimensional protein index towards a response-

RT regulation map for Bacillus subtilis";

RL Electrophoresis 18:1451-1463 (1997).

CC -! CATALYTIC ACTIVITY: L-LEUCINE + 2-OXOGLUTARATE = 4-METHYL-2-

CC OXOPENTANOATE + L-GLUTAMATE (ALSO ACTS ON L-ISOLEUCINE AND

CC L-VALINE).

CC -! COFACTOR: PYRIDOXAL PHOSPHATE.

CC -! PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS.

CC -! SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT

CC AMINOTRANSFERASES.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; Z49992; CAA90289.1; -

DR EMBL; X73124; CAA51556.1; -

DR EMBL; Z99123; CABI5881.1; -

DR PIR; S39655; S39655.

DR HSP; P00510; 1A3G.

DR SubtilList; BG10546; ywaa.

DR InterPro; IPR001544; Aminotran\_4.

DR Pfam; PF01063; aminotran\_4; 1.

DR ProDom; PD001961; Aminotran\_4; 1.

DR PROSITE; PS00770; AA\_TRANSFER\_CLASS\_4; 1.

KW Transaminase; Aminotransferase; Branched-chain amino acid biosynthesis;

KW Pyridoxal phosphate; Complete proteome.

FT INIT\_MET 0

FT BINDING 196 196 PYRIDOXAL PHOSPHATE (BY SIMILARITY).

SQ SEQUENCE 362 AA; 40191 MW; FFA681A5FDC7057 CRC64;

alignment\_scores:

Quality: 104.50 Length: 328

Ratio: 0.661 Gaps: 15

Percent Similarity: 48.171 Percent Identity: 22.561

alignment\_block:

US-09-303-518D-131 x ILVE\_BACSU ..

Align seg 1/1 to: ILVE\_BACSU from: 1 to: 362

```
160 GCCAAGTCGCTGTTGAAGACAAAAG.....AATCCGGG 194
|||||:|||||:|||||:|||||:|||||:
66 GlyGlnThrValPheGluGlyLeuLysAlaTyrValSerGluAspAsp 82
195 CGTAGTATTACTGCGCGCTTCAGGCAAAATCGCGCTATTACCGGTG 244
|||||:|||||:|||||:|||||:|||||:
82 svalLeuLeuPheArgPro.....GluLysA 91
245 GCGAAAGCGCTACTTCAGTCAGTCGTGATTGCCGTGAAGCGCAACGAC 294
:::|||||:|||||:|||||:|||||:
91 snMetGluArgLeuAsnGlnSer.....AsnAsp 100
295 GAATCGAGTTC.....GAAGCTACGTACCTGAAGCGCTGGC 332
|||||:|||||:|||||:|||||:|||||:
101 ArgLeuCysIleProGlnIleAspGluGlnValLeuGluGlyLeu 117
333 AAAATTGAGCAGCGAAAGTGCAGCACTGATTCAATCAGGCTTAT 382
|||||:|||||:|||||:|||||:|||||:
117 sGlnLeuValAla.....IleAspLysAspTrpIleProAsnAlaGlu 132
383 GGACTGCGCTTGGCACCGCTCGTTACGCAAAATCCCTGCGGTAGATGCC 432
|||||:|||||:|||||:|||||:|||||:
132 lYThrSerLeuTyrIleArgProPhe.....IleIleAlaThr 144
433 GAGCGTTCGCCATCTTCGTCATTCGATGCACACCAATCCGCTGGCTGC 482
|||||:|||||:|||||:|||||:|||||:
145 GluProPhe...LeuGlyValAlaAlaSerHisThrTyrLysLeu.... 158
483 CGACCTACGCTCATCATCAAGAACGCGGAGAGCTTCAAGCGCGGCC 532
|||||:|||||:|||||:|||||:|||||:
159 .....LeuIleLeuSerProValGlySerTyrTyrLysGluGly 173
533 TGCTGGTATTGAGCGCTGACCGCAACCTAAATC..... 567
:::|||||:|||||:|||||:|||||:
173 lElysProValLysIleAlaValGluSerGluPheValArgAlaValLys 189
568 .....CATGCTGTAAAGCAGCGCAGACGTGCGCTGTGAATA 608
:::|||||:|||||:|||||:|||||:
190 GlyGlyThrGlyAsnAlaLysThrAlaGlyAsnTyrAlaSerSerLeu 206
609 TGCTGCCAATATCGAAACACATGAATTTGGGCGCGCGCATCTGCGGCT 658
|||||:|||||:|||||:|||||:|||||:
206 salAlaGlnVal...AlaGluGluGlyGlyPheSerGlnValLeuTrpL 222
659 TGAGTGGCAGCAGCATTCATTTCATCGAGCGCAGTCGCGCGCAATAAAC 708
|||||:|||||:|||||:|||||:|||||:
222 euAspGlyIleGluLysLysTyrIleGluGluValGlySerMetAsnIle 238
709 GTGTGGACCATCAATATCAAGACGTGATTGCTATCGGACGTTTCTCGT 758
:::|||||:|||||:|||||:|||||:
239 PhePheLysIleAsnGlyGlu.....IleVal 247
759 AACAGCGCGCTCTGAAT..... 774
|||||:|||||:
247 lThrProMetLeuAsnGlySerIleLeuGluGlyIleThrArgAsnSer 264
775 .....ACGAGCGCGTGTGT 789
264 alIleAlaLeuLeuLysHisTrpGlyLeuGlnValSerGluArgLysIle 280
790 GCCTTGGGCGGC...CTGCAAGTCAACAAACCGCGCTCTTTCGTCACGT 836
```

UNLINED (GLCNAc. :.) (POTENTIAL

[illegible][illegible]

FT	1	25	POTENTIAL.
FT	26	676	MUCIN 1.
FT	26	582	EXTRACELLULAR (POTENTIAL).
FT	583	603	POTENTIAL.
FT	604	676	CYTOPLASMIC (POTENTIAL).
FT	458	573	SEA.
FT	291	291	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	323	323	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL).

```
184 yserSerAlaProProThrSerSerAlaValAsnSerAlaThrThrProV 201
607 TTTCAGCGGACGCTGCTGGCTGCTGCTTACACACATGATTTACGT 558
201 alHisSerGlySerAlaPro..... 208
557 TCGGTACGGCGGCTCAATACCAAGCGCGTGTGAAGCTTCTGGCGGC 508
209 .....ValThrSerSerAlaVal 214
507 TTCTTTGATGATGACC.....GTAGGTCGGCAGCC.....A 476
214 lAsnSerAlaThrThrProValHisSerGlySerSerAlaProValThrS 231
475 GCGGATTTGTCCTGCGATTCAGTTCAGAGATGGCGAAGCGGTTCGCATCT 426
231 erSerAlaValAsnSerAlaThrThrProValHisSerGlySerSerAla 247
425 .....ACGGCAGGATTTGCTGTAAGCGG.....GTCCGAGCGC 388
248 ProProThrSerSerValValAsnSerAlaThrThrProValHisSerGI 264
387 AGTCCATAAGCCTGATGATGATGATGATGATGATGATGATGATGATG 338
264 yserSerAlaProPro.....ThrSerSerAlaValA 275
337 ATTTTCCCGCGCTTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 288
275 snLeuAlaThrThrPro.....ValHisSerGlySerSerThr 287
287 CCTTCAACGGCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 238
288 ProAlaThrAsnSerThrThrAspSerAlaThr.....ThrPro..... 300
237 AATAGCGGGGATTTTCCTGAGCGCGCGCGCGCGCGCGCGCGCGCGG 188
301 .....ValProGlySerSerMetGlnThrThrGluAlaI 313
187 TCCTTTTGTCTTCAAC.....AGCATTGGCGCTTTT 156
313 leSerGlySerAlaAsnThrProIleHisAsnGlySerLeuValPro... 328
155 TTGACGGCTTCACCTTCCTGATTTTCATCGAGGGCGCGATCGCGACATA 106
329 ...ThrThrSerSerAlaLeuVal.....ProThrThr 338
105 TTCTTCCCGCAGCAGCGCTTCGTTGTAATGCGCGCGCGCGCTCA 63
338 rSerAlaAlaHisSerGlyAlaSerAlaMetThrAsnSerSer 352
seq_name: SwissProt_40:MURF_HAEIN

seq_documentation_block:
ID MURF_HAEIN STANDARD; PRT; 457 AA.
AC P45061;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAC-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
OS MURF OR Hil134.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

```

[illegible]

```

345 yValcyl.....GlnSerGluGluAsnValArg 354
157 RAAGCCAAAGTCTGTTGAAGCAAAAAGATCCGGCGTAGTATTAC 206
    ||||| ||| :|:|:|
355 Lys.....LeuPheAlaAspAlaGluArg 362
207 TCGCGCGCTTCAGGCAAAATCGCGCTATTACCGCTGGCGAAAGCCG 256
    ||||| ||| :|:|:|
363 .....GluTyrArgAspArgGlyGluGluSerG 372
257 TACTTCAGTCAGTCGTGATT..... 276
    ||||| :|:|:|
372 LysLeuHisLeilellePheAspGluLeuAspAlaIleCysLysLysArg 388
277 GCGGTGTAAGCAAGCAAGCAAAATCGAGTTCGAACGCTACGTACCTGAAGC 326
    :|:|:| ||| :|:|:|
389 GlySerSerGlyGlyAspThrGlyValGlyAspGlnValValAsnGlnLe 405
327 GCTGGCAAAATGAGCAGCAAGAAAGTCGCCGCAACCTGATTCATCAG 376
    ||||| ||| :|:|:|
405 uLeuAlaLysMetAspGlyValAspGlnLeuAsnAsnIleLeuValIleG 422
377 GCTTATGAGACTCGGCTTCGCAACCGCTCCGTTCCAGCAAAATCCCTCCGCTA 426
    |||:|
422 LysMet..... 423
427 GATGCGGAGCGGTCCGCAATCTGTCATGCGATGCGACCAATCCGCT 476
    :|:|:| |||
424 .....ThrAsnArgLysAsp..... 428
477 GGCTGCGGACCTACGCTCATCAAGCAAGCGCGCAAGACTTCAAC 526
    :|:|:| ||| |||||
429 .....MetIleAspGluAlaLeu.....LeuArgp 437
527 GCGCGCTGTGTGATTG.....ACC 546
    ||| ||| :|:|
437 roglyArgLeuGluValHisMetGluIleSerLeuProAspGluHisGly 453
547 CGCTGACCGCAAGCAAGTAAATCCATGTCGTAAGCAGCAGCGC..... 588
    ||||| ||||| |||||
454 ArgLeuGlnIleLeuLysIleHisThrSerArgMetAlaSerAsnGlyIle 470
589 .....GCAGACGTCGCGCTCTGAAATGTCGCAATATCGCAACACATG 631
    ||||| ||||| |||||
470 eLeuGluAsnAspValAspMetGluGluLeuAlaSerLeu...ThrLysA 486
632 AATTGCGCGCGCGATCTCTCGCGCTG.....AGTGGCAGCGCATTT 675
    :|:|:| ||| :|:|:|
486 snPheSerGlyAlaGluIleAlaGlyLeuIleLysSerAlaSerPhe 502
676 CATTTTC..... 681
    |||
503 AlaPheTyrArgHisIleLysValGlyThrThrAlaAlaValSerGlyAs 519
682 .ATGAGCGCGCGCGCGAATAAACCGGTGG..... 714
    :|:|:| ||| :|:|:|
519 nLeuGluAsnIleLysValAsnArgAsnAspPheLeuAsnAlaLeuSerG 536
714 ..... 714
536 luValArgProAlaTyrGlyValSerGluGluLeuGluSerArgVal 552
715 .....ACCATCAATTAT.....CAAGACGTGATTCGTAT 743
    ||||| |||||
553 GlnGlyGlyIlelleAsnPheGlyLysHisIleGluGluIlelleThrGI 569
744 CGGACGTTTGTGTACACGCGCTCTGATACCGGCGC..... 783
    ||||| ||||| |||||
569 uGlyLysLeuPheValGlnGlnValLysAsnSerGluArgThrArgLeuV 586
784 .....GTGGTTGCTTGGCGCGCTGCAAGTCAACAAACCGCGCTCTTG 828
    ||||| ||| :|:|:|
586 alSerValLeuLeuSerGlyProIleAlaSerGlyLysThrAlaLeuAla 602

```

```

829 CGTACCGCTTTGGTGGAGAGGTGCTCAACTTACCGCGCGGAATTGCT 878
    ||||| ||| :|:|:|
603 AlaThrIle.....AlaLeuGlySerGluPhePropheValLysLeuVa 617
879 TGACGCGGACACACCGGTGATTTCCCGTTCGGTATTGAACGCTGCGATTG 928
    :|:|:| ||| ||| :|:|:|
617 lSerAlaGluSerMetVal.....GlyMetAsnGluAsnAlaArgValA 632
929 CACAAGCGCGCATGATTATTTCGGGAGCGGTACCAACATCAGATTTCGGTT 978
    ||||| :|:|:| ||| :|:|:|
632 laHisValAsnArgValPheGluAspSerTyrLysSerProLeuSerVal 648
979 ATCGAAGAAGCGCGCAGCAAGAGAGCTGTCGCTGGTTCGCGCGCAGCGC 1028
    ||| :|:|:| ||| |||||
649 lIleValValAspGluIleGluArgIlelleAspTrpVal...ProIleGI 664
1029 GGCAAAATCTCCATCAGCGCAGCGCTCTCGGCGCATTTCTCTAAAAACA 1078
    :|:|:| ||| ||| :|:|:|
664 yProArgPheSerAsnThr.....LeuLeuGlnThrL 675
1079 AACTCTTCAAGTTCACGACAGCGCTCAACGCGCGCAGCGCGCATGGTA 1128
    ||| ||| ||| |||
675 euMetValLeuPheLysGlnProLysGlyHisArgLeuLeuIle 691
1129 CCGATCGCGCACTTATGAGCGCGTAATG 1155
    ||| ||||| |||||
692 LeuAlaThrThrSerGluArgThrMet 700
seq_name: SwissProt_40:VGP3_EBV

seq_documentation_block:
ID VGP3_EBV STANDARD; PRT; 907 AA.
AC P03200; P03201;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
GN Glycoprotein GP220].
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.; of the B95-8 Epstein-Barr virus genome."
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
RL Nature 310:207-211(1984).
CC -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC B-CELLS.
CC -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC ENVELOPE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01555; CAA24854.1; -.
DR PIR; A03762; Q0BE21.
DR PIR; A03763; Q0BE22.
DR PIR; S33008; S33008.
KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
FT CARBOHYD 47 47 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).

```



```

FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 610 610 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 656 656 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 858 858 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 888 888 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLIC 502 698 MISSING (IN GP220)
SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;

```

```

alignment_scores:
  Quality: 103.50      Length: 507
  Ratio: 0.470         Gaps: 26
  Percent Similarity: 43.393      Percent Identity: 22.485

```

alignment\_block:

US-09-303-518D-131/rev x VGP3\_EBV

Align seg 1/1 to: VGP3\_EBV from: 1 to: 907

```

1340 CCTTCCTCTCAATGGTTTCCAGCACTTTCGCCAAGCGGGCGG..... 1296
||||| : : : : : : : : : : : : : : : : : : : : : : : :
240 ProSerGlyIleLeuThrSerThrSerProValAlaThrProIlePr 256
1295 .....TATTCGTATTTCGCCGGCAGACGAGCTGCACAAAGCGA 1256
||||| : : : : : : : : : : : : : : : : : : : : : : : :
256 oGlyThrGlyTyrAlaTyrSerLeuArgLeuThrProArgProValSera 273
1255 GTCTCTTCCTCCCAATTC..... 1230
|| : : : : : : : : : : : : : : : : : : : : : : : :
273 rGheLeuGlyAsnAsnSerIleLeuTyrValPheTyrSerGlyAsnGly 289
1229 CCCAAAGCCTCGCG..... 1215
||||| : : : : : : : : : : : : : : : : : : : : : : : :
290 ProLysAlaSerGlyAspTyrCysIleGlnSerAsnIleValPheSe 306
1214 .....CTGTCGGTATCCCGCAGGATTAAATCGCGCAAAAGGAGTAG 1172
: : : : : : : : : : : : : : : : : : : : : : : :
306 rAspGluIleProAlaSerGlnAspMetProThrAsnThrThrAspIleT 323
1171 GCAGGATGCCAACGCATTACCGCTCATAGTCCGATCGGTACCATG 1122
: : : : : : : : : : : : : : : : : : : : : : : :
323 hrTyrValGlyAspAsnAlaThrTyrSer...ValProMetValThrSer 338

```

```

1121 .....GCGGGTCCGCGCGTTCAGCGCTGCTGCTG..... 1092
||||| : : : : : : : : : : : : : : : : : : : : : : : :
339 GluAspAlaAsnSerProAsnValThrValThrAlaPheTrpAlaTrpPr 355
1091 ...AAGTTTGAAGAGTTGTTTGGAAATGGCCGAGAGTGGTGGCG 1046
||||| : : : : : : : : : : : : : : : : : : : : : : : :
355 oAsnAsnThrGluThrAspPheLysCysLysTrpThrLeuThrSerGlyT 372
1045 TGATGGAGTATTGTCGGCTGC.....GCGCAACCCAGCGC 1008
: : : : : : : : : : : : : : : : : : : : : : : :
372 hrPro.....SerGlyCysGluAsnIleSerGlyAlaPheAlaSer 385
1007 AACAGCTCTTTGTCGGCCCTTCTTCGATAACGGAATCTGATTGGTA 958
||||| : : : : : : : : : : : : : : : : : : : : : : : :
386 Asn.....ArgThrPheAspIleThr..... 392
957 GCGTCCCAATAATCATGCGCGCTTGTGCAATCGCACCGTTCAATACCG 908
393 .....ValSerGlyLeuGlyThrA 399
907 AACCGGAA.....ATCACGCGTTGTCGCGTCAACCAATTCGCGC 867
||||| : : : : : : : : : : : : : : : : : : : : : : : :
399 laProLysThrLeuIleIleThrArgThrAla..... 409
866 GCGGTAAGTTGACACACCTTCGCACCCAAACGGTACGCAAGGCGCG 817
||||| : : : : : : : : : : : : : : : : : : : : : : : :
410 .....ThrAsnAlaThrThrThrHisLys..... 418
816 TTTGTTGACTTGACGCGCCCAAGCAACCGCTCGGTATTCAGAG 767
||||| : : : : : : : : : : : : : : : : : : : : : : : :
419 ValIlePheSerLysAlaProGluSerThrThrThrSer..... 431
766 GGCTCTTACGAAACACCTCGGATAGCAATCAGTCTTTGATAATGATG 717
||||| : : : : : : : : : : : : : : : : : : : : : : : :
432 ..ProThrLeuAsnThrThrGlyPheAlaAspProAsn..... 443
716 GTCCACACGGTTTATTTCGCGCGCTGCTCGATCAATGAATGTCGT 567
||||| : : : : : : : : : : : : : : : : : : : : : : : :
444 .....ThrThrThrGlyLeuProSerSerThrHisVa 454
666 GCCACTCAAG.....CCGCGAGGATGCGCGCCCAATTCATG 626
||||| : : : : : : : : : : : : : : : : : : : : : : : :
454 lProThrAsnLeuThrAlaProAlaSerThrGlyProThrValSerThrA 471
625 TTTTCGATATGCGCAGCATTTTCAGAGCGCAGCTGCGCTGCTCTTTA 576
||||| : : : : : : : : : : : : : : : : : : : : : : : :
471 laAspValThrSerProThrProAlaGlyThrThrSerGlyAlaSerPro 487
575 CACACA.....TGG..... 567
||||| : : : : : : : : : : : : : : : : : : : : : : : :
488 ValThrProSerProSerProThrAspAsnGlyThrGluSerLysAlaPr 504
566 ....ATTTTACCTTCGCGCAGCGCTCAATACCAACAGCGCGCTTGA 521
: : : : : : : : : : : : : : : : : : : : : : : :
504 oAspMetThrSerSerThrProValThrThrProThrProAsnAlaT 521
520 AGTCTTCGCGGCTTCTTTGATGACCGTAGGTCGCGCAGCCAGCGGA 471
||||| : : : : : : : : : : : : : : : : : : : : : : : :
521 hrSerProThrProAlaValThrThrProThrProAsnAlaThrSerPro 537
470 TTGTTGTCATGCGCATTCAGCAAGATGCGCAACGCGCTCGCATACGCG 421
||||| : : : : : : : : : : : : : : : : : : : : : : : :
538 ThrProAlaValThr...ThrProThrProAsnAlaThrSerProThrLe 553
420 AGGGATTTCGTGAACGCGCGGTGCGAAGCGCAGTCCATGAAGCTGATT 371
||||| : : : : : : : : : : : : : : : : : : : : : : : :
553 uGly.....LysThrSerProThrSerAlaValThrThrProThrPr 567
370 GAATCAGGTTTGGCGCACTTTTTCGCTGCTCAATTTTCCAGCGCTTCA 321
||||| : : : : : : : : : : : : : : : : : : : : : : : :
567 roAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaVal 583
320 GGTACGTAGCGCTTCGAACCTCGATTTCG.....TCGTTGCC 286

```



```

695 GCGCGAATAAAACCGTGTGGACCATCAATATCAAGACGTGATTCCTATC 744
||||| : : : : : ||||| : : : : :
367 roAlaAsnThrSerAlaTrpThrLeuHis.....AlaAla 378
745 GGA..... 747
|||
379 GlyThrGluSerGlyAlaAsnAlaAlaThrAlaThrAlaProSerPheAs 395
748 ..CGTTGTTGCTGAACAGCGCTCTGAATACAGCGGTGTTCCCTGG 796
: : : : : ||||| : : : : :
395 pGluAlaPheLeuThrAspArgLeuGlnLeuLeuLeuHisAla.... 410
||| : : : : :
797 GCGCGCTGCAAGTCAACAAACCGCGCTCTTTCGCTACCGTTTGGTGGC 846
||||| : : : : :
411 .....ValAsnGlnArgSerCysLeuArgArgProCysGlyPro 423
847 AAGGTGCTCAACTACCGCGCGGAATGTTGACGCGGACACCGCGT 896
: : : : : ||| : : : : :
424 GlnSerAlaAlaGlnGlnAlaValArgAlaTyrLeuSerLysLy 440
||| : : : : :
897 GATTTCGGTTCGGTATTGAAC.....GNGCGGATTGCCAAG 934
: : : : : |||||
440 sLeuAspAlaPheLeuLeuAsnTrpLeuHisHisGlyLeuAspLeuGln 457
935 GCGCGCATGATTATG..... 951
|||||
457 r9MetHisAspTyrLeuSerHisLysThrThrLysGlyThrYrSerThr 473
952 ..... 953
GG
474 LeuAspArgAlaLeuLeuGluLysMetGlnValValPheAspProTyrG 490
954 ACGTATCAC.....AATCAGATTCCG 976
|||||
490 yArgGlnHisGlyProAlaLeuLeuAlaTrpValGluGluMetLeuArg 507
977 TATCGAAGAGCGCGCAGCAGACAGCTGTCGGCTGGGTGGCGCGCAG 1026
: : : : : |||||
507 yValGluSerLysProThrAsnGluLeu.SerGlnArgLeuGlnArgPh 523
1027 CCGGACAAATACTCCA.....TCACGCGCACCACTCTCGG 1061
|||||
523 eValThrLysArgProMetProValSerAspSerPheValCysLeu.Arg 539
1062 CCATTTCTCTAAACAAACTCTTCAAGTTCAGGACGACGCGGTCAACGG... 1109
||| : : : : :
540 ProValAspPheGlnArgLeuThrGlnValIleGluGlnArgArgVa 556
1110 .....CGCGCAGCGCGCATGGTACCGATCGCACTTATGAG 1146
|||||
556 lLeuGlnArgGlnArgGluGluTyrHisGlyValTyrGluHisLeu.... 571
1147 CGCGTAATGCGGTGGACATCTCGCTACCTTGTCTTTTCGCGGATTAAT 1196
: : : : : |||||
572 .....AlaGly.LeuIleThrSerIleAspIleHisAspLeuAs 584
1197 CGTGGCGATACGACGACGCGCGCTTGGTGGTGGTGGTGGTGGTGG 1243
: : : : : |||
584 pAlaSerAspLeuAsnArgGluIleLeuLysAlaLeuGlnProLeuA 601
1244 ACGAA 1248
|||
601 spAsp 602

```

seq\_name: SwissProt\_40:SGS3\_DROER

seq\_documentation\_block:

```

ID SGS3_DROER STANDARD; PRT; 328 AA.
AC P13730;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Salivary glue protein Sgs-3 precursor.

```

```

GN SGS3.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88332966; PubMed=3138416;
RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;
RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";
RL J. Mol. Biol. 201:273-287(1988).
CC -!- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: M14652; -; NOT_ANNOTATED_CDS.
DR PIR: S01359; S01359.
DR FlyBase: FBgn0012268; Dere\SGs3.
KW Repeat; Signal.
FT SIGNAL 1 23
FT CHAIN 24 328
FT SEQUENCE 328 AA; 36355 MW; 62F27F188C0F8272 CRC64;
SQ SEQUENCE 328 AA; 36355 MW; 62F27F188C0F8272 CRC64;

```

alignment\_scores:

```

Quality: 102.00 Length: 399
Ratio: 0.729 Gaps: 14
Percent similarity: 35.088 Percent Identity: 20.050

```

alignment\_block:

US-09-303-518D-131 x SGS3\_DROER ..

Align seg 1/1 to: SGS3\_DROER from: 1 to: 328

```

29 TGCCCATCGCGGCGAGACCGGAGCAAGTCATTTATGACGCGCCGG ..... 73
|||||: : : : : |||||
28 CysProLysArgThr...ThrProLysProCysThrThrAlaArgProTh 43
74 .....CCATYACCGAAGTCGGTTCGTCGCGAAGAATATGTCGGCAT 116
||||| : : : |||
43 rCysAlaProValThrThrThrCysArgProProThrThrThrArgC 60
117 GCGCCCTCGATGAAATCAAGGAAGGTCAAGCCGTCAGAAAGGCCCAAG 166
|| ||| : : : |||
60 ySProProProThrThrThrArgCysProProProProThrArgProAlaGlu 76
167 TGCTGTTTGAAGACAAAGAAATCCGGCGTAGTATTACTGCGCGCGCT 216
||| ||||| |||
77 CysThrAlaThrThrLysArgProThrAlaArgProThrThrArgArgTh 93
217 TCAGGCAAAATCGCGCTATTCCCGTGGCGGAAAGCGCGTACTTCAGTC 266
: : : : :
93 rThrValArgAla..... 97
267 AGTCGTGATTGCGGTGAAGCAACGAGCAATCGAGTTCGAAACGCTACG 316
: : : ||||| : : :
98 .....ThrThrLysArgAlaThrThrArgArgThrThrLysArgAla 111
317 TACCTCAAGCGCTGCAGAAATGAGCAGCGAAAGTGCAGCGCAACCTG 366
: : : |||
112 ThrThrArgArgThrThrValArgAlaThrThrLysArgAlaThrThr.. 127
367 ATTCAATCAGCGCTTATGGACTGCGCTCGCACCGCTCCGTTCCAGCAAAAT 416
: : : : : ||||| : : :
128 .....ArgArgThrThrThrLysArgAlaProThrArgArgThrThr 442

```

```

417 CCCTGCGGTAGATGCCGAGCGGTGCGCACTTCGTCAATGCGATGACGACA 466
   ::::::::::::::::::::
142 hrLysArgAlaThrThrArgArgAsnProThr.....ArgArgThr 155
   ::::::::::::::::::::
467 CCAATCGCTGGTGGCGGACCGCTACGCTCATCAAGAGAGCGCGGAA 516
   ::::::::::::::::::::
156 ThrThrArgAlaProThrLysArgAlaThrThrLysArgAlaThrTh 172
   ::::::::::::::::::::
517 GACITCAACGCGCTGTTGGTATTGAGCGCGCTGACCGAAGCTAAAT 566
   ::::::::::::::::::::
172 rArgArgAsnProThr..... 177
   ::::::::::::::::::::
567 CCATGTGTAAAGACAGCGCGAGCGCTGCGCTGTAATGCTGCCA 616
   ::::::::::::::::::::
178 .....LysArgLysThrThrArg..... 183
   ::::::::::::::::::::
617 ATATCGAACACATGAATTTGGCGCGCGCATCTGCGCGCTTGTGTCG 666
   ::::::::::::::::::::
183 ..... 188
   ::::::::::::::::::::
667 AGCGACATTCATTCATCGAGCGCGAGTGGCGCGAATAAAACCGTGTG 716
   ::::::::::::::::::::
184 .ArgThrThrVal..... 187
   ::::::::::::::::::::
717 CATCAATTATCAAGACGTGTTGCTATCGGACGTTTGTCTACACGCG 766
   ::::::::::::::::::::
187 ..... 187
   ::::::::::::::::::::
767 GTCTGAATACCGAGCGCGTGTGCTGCTGGCGCGCTGCAAGTCAACAA 816
   ::::::::::::::::::::
188 .....ArgAlaThrLysThrLys 194
   ::::::::::::::::::::
817 CGCGCGCTCTGCTACCGTGTGTTGGTGGAGAGTGTCTCAACTACGC 866
   ::::::::::::::::::::
194 sArgAla..... 196
   ::::::::::::::::::::
867 CGCGCAATTTGTTGACGCGGACCGCGTATTTCGGTTCGTTATGTA 916
   ::::::::::::::::::::
197 .....ThrThrLysArgAlaProThrLys..... 204
   ::::::::::::::::::::
917 ACGGTGCGATTGACAAAGCGCGCATGATTATTTGGGACGTACACAA 966
   ::::::::::::::::::::
205 .....ArgAlaThrThrLys 209
   ::::::::::::::::::::
967 CAGATTTCG.....TTATCGAGAGCGCGCACCAAGAGCT 1004
   ::::::::::::::::::::
209 sArgAlaProThrLysArgAlaThrThrLysArgAlaProThrLys 226
   ::::::::::::::::::::
1005 GTTCGCTGGTGGCGCGCGCGGACAAATATCTCCATCGCGCACCA 1054
   ::::::::::::::::::::
226 laThrThrLysArgAlaProThrLysArgAlaThrThrLysArgAla 242
   ::::::::::::::::::::
1055 CTCTCGCGCATTCCTTAAACAACTCTCAAGTTCACGACCGCTC 1104
   ::::::::::::::::::::
243 ThrLys.....ArgAlaThrThrLysArgAlaProThr 253
   ::::::::::::::::::::
1105 AACGCGCGGACG.....GCGCATGTTACGATCGGCACTATGA 1145
   ::::::::::::::::::::
253 rLysArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLys 270
   ::::::::::::::::::::
1146 GCGGTAATCGCTGGACATCC.....TGCTACTGTC 1180
   ::::::::::::::::::::
270 rGAlaThrAlaArgProThrSerLysProCysGlyCysLysProCys 285
   ::::::::::::::::::::
seq_name: SwissProt_40:ATPB_NICPL
seq_documentation_block:
ID ATPB_NICPL STANDARD; PRT; 498 AA.
AC P26529;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP synthase beta chain (EC 3.6.3.14).

```

```

GN ATPB.
OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco), and
OS Nicotiana bigelovii (Bigelow's tobacco).
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092, 4088;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=92390714; PubMed=1387730;
RX Avni A., Anderson J.D., Holland N., Rochaix J.-D., Gromet-Elhanan Z.,
RA Edelman M.;
RT "tentoxin sensitivity of chloroplasts determined by codon 83 of beta
RT subunit of proton-ATPase.";
RL Science 257:1245-1247(1992).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC H(+)(Out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X61320; CAA43613.1; -.
CC EMBL; X61316; CAA43609.1; -.
CC PIR; S15723; PWNTBC.
CC PIR; S15722; PWNTBB.
CC HSSP; P00829; IBMF.
CC Mendel; 2235; NICbi:atpB.1.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR004100; ATP-synt_ab_N.
CC InterPro; IPR000793; ATPase_AB_C.
CC InterPro; IPR000194; ATPase_alpha_beta.
CC Pfam; PF00006; ATP-synt_ab; 1.
CC Pfam; PF00306; ATP-synt_ab_C; 1.
CC Pfam; PF02874; ATP-synt_ab_N; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis; Chloroplast; Thylakoid; Membrane; CF(1);
KW Hydrolase; ATP-binding; Hydrogen ion transport.
FT NP_BIND 172 179 ATP (BY SIMILARITY).
SQ SEQUENCE 498 AA; 53525 MW; 94ADC63E8BCBF5B8 CRC64;

```

alignment\_scores:  
Quality: 102.00 Length: 492  
Ratio: 0.493 Gaps: 28  
Percent Similarity: 42.073 Percent Identity: 22.154

alignment\_block:

US-09-303-518D-131 x ATPB\_NICPL ..  
Align seq 1/1 to: ATPB\_NICPL from: 1 to: 498

```

178 GACAAAAAGAAATCCGGC.....GTAGTATT 203
   ::::::::::::::::::::
16 GluLysLysAsnProGlyArgValValGlnIleIleGlyProValLeuAs 32
   ::::::::::::::::::::
204 TACTGCGCGCGCTTCAGGCAAAATC.....GCCGTATTC 238
   ::::::::::::::::::::

```

```

32 pValAlaPheProProGlyLysMetProAsnIleTyrAsnAlaLeuValV 49
239 ACCGTGGCAAAAGCGGCTTCACTCAGTCAGTCGATGCGGTGAA... 285
   :::::  |||  |||  :::::  |||  |||
49 alGlnGlyArgAspSerValGlyGlnProIleAsnValAlaCysGluVal 65
286 .....GGCAACGAGCAAAATCGAGTTCGACGCTACGTACCTGA 323
   |||||  :::::  :::::  :::::  :::::  :::::  Ar 75
66 GlnGlnLeuLeuGlyAsnAsnArgVal.....Ar 75
324 AGCGCTGGCAAAATGAGCAGCAAAAGTCCGCCGCAACCTG...A 367
   :::::  |||  |||  :::::  :::::  |||  |||
75 gAlaValAlaMetSerAlaIleAspGlyLeuThrArgGlyMetGluValI 92
368 TTCAATCAGGCTTATGAGTCTCGCTTCGCACCCGCTCGCTTCAGCAAAATC 417
   |||||  |||  |||  |||  |||  |||  |||
92 LeAspThrGly.....AlaProIleSer...Val 100
418 CCTGCCGTAGATGCCGAGCGCTTCGCCATCTTC..... 450
   |||||  |||  |||  |||  |||
101 ProValGlyAlaIleThrLeuGlyArgIlePheAsnValLeuGlyGluPr 117
451 .....GTCAATGCGATGGACACCAATCCGCTG... 477
117 oValAspAsnLeuGlyProValAspThrSerThrThrSerProIleHisA 134
478 ..GTCGCGGACCCCTACGCTCATCATCAAGAACGCCGCGCAAGACTTCAAA 525
   :::::  |||  |||  |||  |||  |||
134 rgSerAlaProAlaPheIleGlnLeuAspThrLysLeuSerIlePheGlu 150
526 CGCGGCTGTGTGATTCAGCCGCTCAGCAGCAAGTAAATCCTATGCTGTG 575
   |||||  |||  |||  |||  |||
151 ThrGlyIleLysValValAspLeuLeu..... 159
576 TAAAGCAGCAGCGCAGACGCTGCGCTGCTGAAATGCTGCCAATATCGAAA 625
   :::::  |||  |||  |||  |||
160 .....AlaProTyrArgArgGlyGlyIleGlyL 170
626 CACATGAATTTGGGCGCCGCTCCTCGCGCTTGTAGTGCACGACATT 675
   |||||  |||  |||  |||
170 eu.....PheGlyGly..... 173
676 CATTTTCATCAGCGCAGCGCGGATTAACCGTGTGGACCATCAATTA 725
174 .....AlaGlyValGlyLysThrValLeuIleMetGluLe 185
726 TCAAGACGTGATGCT.....A 742
   :::::  |||  |||  |||
185 uIleAsnAsnIleAlaLysAlaHisGlyGlyValSerValPheGlyGlyV 202
743 TCGGACGT.....TTGTTCTGTA..... 759
202 alGlyGluArgThrArgGluGlyAsnAspLeuTyrMetGluMetLysGlu 218
760 ACAGGCGCTCTGAATACCGAGCGCGTG.....GTTCCCTGGG 797
219 SerGlyValIleAsnGluGluAsnIleAlaGluSerLysValAlaLeuVa 235
798 CGGCTGCAAGTCAACAAACCGCGCTCTTGTGCTACCGTTTGGGTGCGGA 847
   |||||  |||  |||  |||
235 lTyrGlyGlnMetAsnGluProProGlyAlaArgMetArgValGly... 250
848 AGGTGTCTCACTACCGCGCGGCA..... 873
251 ..LeuThrAlaLeuThrMetAlaGluTyrPheArgAspValAsnGluGln 266
874 .....TTGTTTACCGCGGCAAC.....CGGTGATTTCCGGTTC 908
   |||||  |||  |||  |||
267 AspValLeuLeuPheIleAspAsnIlePheArgPheValIleAlaGlySe 283
909 GGTATTGAACGGTGCATGTCACAA.....GGCGGCGCATGATTAT... 948
   :::::  |||  |||  |||
283 rGluValSerAlaLeuLeuGlyArgMetProSerAlaValGlyTyrGlnP 300

```

```

949 .....TTGGGACGCTACCAACATCAGATTTCCGTTATC 981
   :::::  |||  |||  :::::  |||  |||
300 roThrLeuSerThrGluMetGlySerLeuGlnGluArgIleThrSerThr 316
982 GAAGAAGCGCGCAGCAAGAGCTTTCGGCTGGGTGGCCGCCGCGCAG 1031
   ::|||  |||  |||  |||  |||  |||
317 LysGluGlySerIleThrSerIleGlnAlaValTyrValProAlaAspAs 333
1032 CAATFACTCC...ATCACCGCAGCAGCTCTCGGCCATTTCTTAAAAACA 1078
   |||  :::::  |||||  :::::  |||
333 pLeuThrAspProAlaProAlaThrThrPheAlaHis..... 345
1079 AACTCTTCAAGTTACGACAGCGCTCAACGCGCGCGCATGCGTGGTA 1128
   :::::  |||||  |||  |||  |||  |||
346 .....LeuAspAlaThrThrValLeuSer.....ArgGlyLeuAla 357
1129 CGATCGGCCTTATGAGCGCGTAATGCGTTGGACATCCTCGCTACCTT 1178
   |||  |||  |||  |||  |||  |||
358 AlaLysGlyIleTyrProAlaValAspProLeuAspSerThrSerThr... 373
1179 GCTTTTGGCGGATTAATCGTCGGCGAT..... 1206
   ::|||  |||||  |||  |||  |||
374 .MetLeuGlnProArgIleValGlyGluGluHisTyrGluThrAlaGlnA 390
1207 .....ACCGACAGCGCGCAG 1221
390 rgValLysGlnThrLeuGlnArgTyrLysGluLeuGlnAspIleIleAla 406
1222 GCTTTGGTGTGCTGGAATGACGAAGAACCTCGCTTGTGTCAGC... 1269
   |||||  |||||  |||||  |||||  |||
407 IleLeuGlyLeuAspGluSerGluGluAspArgLeuLeuValAlaAa 423
1270 .....TTCGTCTGC..... 1278
423 gAlaArgLysIleGluArgPheLeuSerGlnProPheValAlaGluV 440
1279 .....CCGGGCAAAATAC 1290
   |||||  |||||
440 alPheThrGlySerProGlyLysTyr 448
seq_name: SwissProt_40:HUTH_BACSU
seq_documentation_block:
ID HUTH_BACSU STANDARD; PRT; 508 AA.
AC P10944;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN HUTH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257040; PubMed=2454913;
RA Oda M., Sugishita A., Furukawa K.;
RT "Cloning and nucleotide sequences of histidase and regulatory genes
in the Bacillus subtilis hut operon and positive regulation of the
hut operon.";
RL J. Bacteriol. 170:3199-3205(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
genome containing the hut and wapA loci.";
RL Microbiology 141:337-343(1995).
CC -1- CATALYTIC ACTIVITY: L-histidine -> urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- INDUCTION: BY HISTIDINE.  
CC -!- PTM: CONTAINS AN ACTIVE SITE 4-METHYLIDENE-IMIDAZOLE-5-ONE (MIO),  
CC WHICH IS FORMED AUTOCATALYTICALLY BY CYCLIZATION AND DEHYDRATION  
CC OF RESIDUES ALA-SER-GLY (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE PAL / HISTIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; M20659; AAA22538.1; -;  
DR EMBL; D31856; BAR06644.1; -;  
DR EMBL; Z99124; CABI5971.1; -;  
DR PIR; S18810; UFBSSH.  
DR HSP; P21310; 1B8F.  
DR Subtilist; BG10687; huth.  
DR InterPro; IPR001106; PAL.  
DR Pfam; PF00221; PAL; 1.  
DR PROSITE; PS00488; PAL-HISTIDASE; 1.  
KW Lyase; Histidine metabolism; MODIFIED TO FORM 4-METHYLIDENE-IMIDAZOLE-  
FT SITE 141 143 5-ONE (BY SIMILARITY).  
SQ SEQUENCE 508 AA; 55674 MW; 869C323BFCC318E0 CRC64;

alignment\_scores:  
Quality: 101.50 Length: 320  
Ratio: 0.668 Gaps: 18  
Percent Similarity: 47.500 Percent Identity: 23.750

alignment\_block:

US-09-303-518d-131 x HUTH\_BACSU ..  
Align seg 1/1 to: HUTH\_BACSU from: 1 to: 508  
22 CTAATCTGCCATCGCGGCGGACACCGGACAGTCAATTTATGACGCG... 69  
151 LeuAlaLeuAlaLeuLeuGlyGlnGlyGlu...ValPhePheGluGlyG1 166  
70 .....CCGGCCATTACCGAGTCGCGTGTTCGGGAGAGATATGTCG 112  
166 uArgMetProAlaMetThr.....GlyLeuLysLysAlaG 178  
113 GCATGCGCCCG...TCGATGAAATCAAGAGGTGAAGCGGTCAAAAAA 159  
178 LysLeuGlnProValThrLeuThrSerLysGluGlyLeuAlaLeuAsn 194  
160 GGCCAAAGTGTGTTGAAGACAAAAAATCCGGCGGTACTATTACTGC 209  
195 Gly.....ThrGlnAlaMe 199  
210 GCCGGCTTACGCAAAATCGCGCTATTACCGTGGGAAAAGCGGTAC 259  
199 tThrAlaMetGlyValValAlaTyrIle...GluAlaGluLysLeuAla 215  
260 TTCAGTCAGTCGTTATGCGC.....GTTGAAGCG..... 288  
215 yrGlnThrGluArgIleAlaSerLeuThrIleGluGlyLeuGlnGly 231  
289 .....AACGAGAAATCGAGTTCGAAGCTACGTACCTGAAGC 326  
232 IleAspAlaPheAspGluAspIleHisLeuAlaArgGlyTyrGlnGlu 248  
327 GCGGCAAAATTCAGCAGCAAAAGTCGCGCGCAACCTGATTCAATCAG 376  
248 nileAsp.....ValAlaGluArgIleArgPheTyrLeuSerAspSerG 263  
377 GCTTATGACTGCG.....CTTCGACCCCGT.....CCGTTTC 408

263 lyLeuThrThrSerGlnGlyGluLeuArgValGlnAspAlaTyrSerLeu 279  
409 AGCAAAATCCCTCGCGGTAGATGCC..... 432  
280 ArgCysIleProGlnValHisGlyAlaThrTrpGlnThrLeuGlyTyrVa 296  
433 ....GAGCGTTCGCATCTTCGTCAATGCGATGACACCAATCCGCTGG 478  
296 llyGluLysLeuGluLleGluMetAsnAlaAlaThrAspAsnProLeu 313  
479 CT.....GCCGACCTACGTCATC..... 498  
313 lPheAsnAspGlyAspLysValIleSerGlyGlyAsnPheHisGlyGln 329  
499 ...ATCAAGAGCGCGCGGACAGACTTCAACGCGCGCTGTTGGTATTGAG 545  
330 ProfileAlaPheAlaMetAspPheLeuLysIleAlaIleSerGluLeuAl 346  
546 CGCGCTGACCGAAGCTAAATCCATGTGTGTAAGCAGCAGCGCGCAGACG 595  
346 aAsnIleAlaGluArgArgIle..... 353  
596 TGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAA.....TTT 636  
354 .....GluArgLeuValAsnProGlnLeuAsnAspLeuProPhe 367  
637 GCGCGCGCGCATCTCGCGGTGAGTGGCAGCGCACATT..... 675  
368 LeuSerProHisProGlyLeuGlnSerGlyAlaMetIleMetGlnTyrAl 384  
676 .....CATTTTCATCGAGCGCGCTGCGCG 697  
384 alaAlaSerLeuValSerGluAsnLysThrLeuAlaHisProAlaSerV 401  
698 CGAATAAAACCGTGTGGACCATCAATTTATCAAGCGGTGTTCTATCGGA 747  
401 alAspSerIleProSerSerAlaAsnGlnGluAspHisValSerMetGly 417  
748 CGTTTGTTCGTAACAGCGCGT.....CTGAATACCGAGCGCGT 785  
418 ThrIleAlaAlaArgHisAlaTyrGlnValIleAlaAsnThrArgArgVa 434  
786 GGTTCGCTTG 795  
434 lIleAlaIle 437